



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12N 15/49, A61K 48/00</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/39302</b> <b>(43) International Publication Date:</b> 6 July 2000 (06.07.00)
<b>(21) International Application Number:</b> PCT/US99/31245 <b>(22) International Filing Date:</b> 30 December 1999 (30.12.99)  <b>(30) Priority Data:</b> 60/114,495      31 December 1998 (31.12.98)    US 60/168,471      1 December 1999 (01.12.99)      US  <b>(71) Applicant:</b> CHIRON CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608 (US).		<b>(72) Inventors:</b> BARNETT, Susan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). ZUR MEGEDE, Jan; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SRIVASTAVA, Indresh; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIAN, Ying; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). HARTOG, Karin; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). LIU, Hong; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). GREER, Catherine; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). SELBY, Mark; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US). WALKER, Christopher; Chiron Corporation, 4560 Horton Street - R440, Emeryville, CA 94608 (US).  <b>(74) Agents:</b> DOLLARD, Anne, S.; Chiron Corporation, Intellectual Property - R440, P.O. Box 8097, Emeryville, CA 94662-8097 (US) et al.  <b>(81) Designated States:</b> AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION OF VIRUS-LIKE PARTICLES		
<b>(57) Abstract</b> <p>The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types, including, but not limited to, mammalian, insect, and plant cells. Synthetic expression cassettes encoding the HIV Gag-containing polypeptides are described, as are uses of the expression cassettes in applications including DNA immunization, generation of packaging cell lines, and production of Env-, tat- or Gag-containing proteins. The invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs including, but not limited to, vehicles for the presentation of antigens and stimulation of immune response in subjects to whom the VLPs are administered.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						



IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND  
PRODUCTION OF VIRUS-LIKE PARTICLES

5     **TECHNICAL FIELD**

Synthetic expression cassettes encoding the HIV polypeptides (e.g., Gag-, pol-, prot-, reverse transcriptase, Env- or tat-containing polypeptides) are described, as are uses of the expression cassettes. The present invention relates to the efficient expression of HIV polypeptides in a variety of cell types. Further, the invention provides methods of producing Virus-Like Particles (VLPs), as well as, uses of the VLPs and high level expression of oligomeric envelope proteins.

15

**BACKGROUND OF THE INVENTION**

Acquired immune deficiency syndrome (AIDS) is recognized as one of the greatest health threats facing modern medicine. There is, as yet, no cure for this disease.

20

In 1983-1984, three groups independently identified the suspected etiological agent of AIDS. See, e.g., Barre-Sinoussi et al. (1983) Science 220:868-871; Montagnier et al., in Human T-Cell Leukemia Viruses (Gallo, Essex & Gross, eds., 1984); Vilmer et al. (1984) The Lancet 1:753; Popovic et al. (1984) Science 224:497-500; Levy et al. (1984) Science 225:840-842. These isolates were variously called lymphadenopathy-associated virus (LAV), human T-cell lymphotropic virus

25

type III (HTLV-III), or AIDS-associated retrovirus (ARV). All of these isolates are strains of the same virus, and were later collectively named Human Immunodeficiency Virus (HIV). With the isolation of a related

5 AIDS-causing virus, the strains originally called HIV are now termed HIV-1 and the related virus is called HIV-2. See, e.g., Guyader et al. (1987) *Nature* 326:662-669; Brun-Vezinet et al. (1986) *Science* 233:343-346; Clavel et al. (1986) *Nature* 324:691-695.

10 A great deal of information has been gathered about the HIV virus, however, to date an effective vaccine has not been identified. Several targets for vaccine development have been examined including the *env*, *Gag*, *pol* and *tat* gene products encoded by HIV.

15 Haas, et al., (*Current Biology* 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (*J. Virol.* 72(2):1497-1503, 1998) described an increased immune  
20 response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage. Schneider, et al., (*J Virol.* 71(7):4892-4903, 1997) discuss inactivation of inhibitory (or instability) elements (INS) located within the coding sequences of the  
25 Gag and Gag-protease coding sequences.

The *Gag* proteins of HIV-1 are necessary for the assembly of virus-like particles. HIV-1 *Gag* proteins are involved in many stages of the life cycle of the virus including, assembly, virion maturation after particle  
30 release, and early post-entry steps in virus replication. The roles of HIV-1 *Gag* proteins are numerous and complex (Freed, E.O., *Virology* 251:1-15, 1998).

Wolf, et al., (PCT International Application, WO 96/30523, published 3 October 1996; European Patent Application, Publication No. 0 449 116 A1, published 2 October 1991) have described the use of altered pr55 Gag of HIV-1 to act as a non-infectious retroviral-like particulate carrier, in particular, for the presentation of immunologically important epitopes. Wang, et al., (Virology 200:524-534, 1994) describe a system to study assembly of HIV Gag- $\beta$ -galactosidase fusion proteins into virions. They describe the construction of sequences encoding HIV Gag- $\beta$ -galactosidase fusion proteins, the expression of such sequences in the presence of HIV Gag proteins, and assembly of these proteins into virus particles.

Recently, Shiver, et al., (PCT International Application, WO 98/34640, published 13 August 1998) described altering HIV-1 (CAM1) Gag coding sequences to produce synthetic DNA molecules encoding HIV Gag and modifications of HIV Gag. The codons of the synthetic molecules were codons preferred by a projected host cell.

The envelope protein of HIV-1 is a glycoprotein of about 160 kD (gp160). During virus infection of the host cell, gp160 is cleaved by host cell proteases to form gp120 and the integral membrane protein, gp41. The gp41 portion is anchored in (and spans) the membrane bilayer of virion, while the gp120 segment protrudes into the surrounding environment. As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells.

Haas, et al., (Current Biology 6(3):315-324, 1996) suggested that selective codon usage by HIV-1 appeared to account for a substantial fraction of the inefficiency of viral protein synthesis. Andre, et al., (J. Virol.

72(2):1497-1503, 1998) described an increased immune response elicited by DNA vaccination employing a synthetic gp120 sequence with optimized codon usage.

5     **SUMMARY OF THE INVENTION**

The present invention relates to improved expression of HIV *Env*-, *tat*-, *pol*-, *prot*-, *reverse transcriptase*, or *Gag*-containing polypeptides and production of virus-like particles.

10       In one embodiment the present invention includes an expression cassette, comprising a polynucleotide encoding an HIV *Gag* polypeptide comprising a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20. In certain embodiments, the polynucleotide  
15       sequence encoding said *Gag* polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9 or SEQ ID NO:4. The expression cassettes may further include a polynucleotide sequence encoding an HIV *protease* polypeptide, for  
20       example a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79. The expression cassettes may further include a polynucleotide sequence encoding an HIV *reverse*  
25       *transcriptase* polypeptide, for example a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ ID NO:84. The expression cassettes may further include a polynucleotide  
30       sequence encoding an HIV *tat* polypeptide, for example a sequence selected from the group consisting of: SEQ ID NO:87, SEQ ID NO:88, and SEQ ID NO:89. The expression cassettes may further include a polynucleotide sequence encoding an HIV *polymerase* polypeptide, for example a

sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. The expression cassettes may include a polynucleotide sequence encoding an HIV polymerase polypeptide, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase. The expression cassettes described above may preserve T-helper cell and CTL epitopes. The expression cassettes may further include a polynucleotide sequence encoding an HCV core polypeptide, for example a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

In another aspect, the invention includes an expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). In certain embodiments, the Env expression cassettes include sequences flanking a V1 region but have a deletion in the V1 region itself, for example the sequence presented as SEQ ID NO:65 (Figure 52, gp160.modUS4.delV1). In certain embodiments, the Env expression cassettes include sequences flanking a V2 region but have a deletion in the V2 region itself, for example the sequences shown in SEQ ID NO:60 (Figure 47); SEQ ID NO:66 (Figure 53); SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:76 (Figure 64) and SEQ ID NO:49 (Figure 36). In certain

embodiments, the Env expression cassettes include sequences flanking a V1/V2 region but have a deletion in the V1/V2 region itself, for example, SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67 (Figure 54); SEQ ID NO:75 (Figure 63); SEQ ID NO:35 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37). The Env-encoding expression cassettes may also include a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide, for example, SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); SEQ ID NO:63 (Figure 50); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may include a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide, for example SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); SEQ ID NO:73 (Figure 61); SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62). The Env expression cassettes may include a gp140 Env polypeptide or a polypeptide derived from a gp140 Env polypeptide, for example SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39

(Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34). The Env expression cassettes may also include a gp120 Env polypeptide or a polypeptide derived from a gp120 Env polypeptide, for example SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42); SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21). The Env expression cassettes may include an Env polypeptide lacking the amino acids corresponding to residues 128 to about 194, relative to strains SF162 or US4, for example, SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the selected host cell. The expression cassettes may be included on one or on multiple vectors and may use the same or different promoters. Exemplary control elements include a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein), a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In another aspect, the invention includes a recombinant expression system for use in a selected host cell, comprising, any one of the expression cassettes described herein operably linked to control elements

compatible with expression in the selected host cell. Exemplary control elements include, but are not limited to, a transcription promoter (e.g., CMV, CMV+intron A, SV40, RSV, HIV-LTR, MMLV-LTR, and metallothionein), a  
5 transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

In yet another aspect, the invention includes a cell  
10 comprising one or more of the expression cassettes described herein operably linked to control elements compatible with expression in the cell. The cell can be, for example, a mammalian cell (e.g., BHK, VERO, HT1080, 293, RD, COS-7, or CHO cells), an insect cell (e.g.,  
15 *Trichoplusia ni* (Tn5) or Sf9), a bacterial cell, a plant cell, a yeast cell, an antigen presenting cell (e.g., primary, immortalized or tumor-derived lymphoid cells such as macrophages, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof).

20 In another aspect, the invention includes methods for producing a polypeptide including HIV *Gag*-, *prot*-, *pol*-, *reverse transcriptase*, *Env*- or *Tat*-containing polypeptide sequences, said method comprising, incubating the cells comprising one or more the expression cassettes  
25 describe herein, under conditions for producing said polypeptide.

In yet another aspect, the invention includes compositions for generating an immunological response, comprising one or more of the expression cassettes  
30 described herein. In certain embodiments, the compositions also include an adjuvant.

In a still further aspect, the invention includes methods of generating an immune response in a subject, comprising introducing a composition comprising one or



more of the expression cassettes described herein into the subject under conditions that are compatible with expression of said expression cassette in the subject.

In certain embodiments, the expression cassette is introduced using a gene delivery vector. More than one expression cassette may be introduced using one or more gene delivery vectors.

In yet another aspect, the invention includes a purified polynucleotide comprising a polynucleotide sequence encoding a polypeptide including an HIV Env polypeptide, wherein the polynucleotide sequence encoding said Env polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59). Further exemplary purified polynucleotide sequences were presented above.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

In another embodiment, the invention includes a method for producing a polypeptide including HIV Gag polypeptide sequences, where the method comprises incubating any of the above cells containing an expression cassette of interest under conditions for producing the polypeptide.

The invention further includes, a method for producing virus-like particles (VLPs) where the method comprises incubating any of the above-described cells containing an expression cassette of interest under conditions for producing VLPs.

In another aspect the invention includes a method for producing a composition of virus-like particles (VLPs) where, any of the above-described cells containing an expression cassette of interest are incubated under

conditions for producing VLPs, and the VLPs are substantially purified to produce a composition of VLPs.

In a further embodiment of the present invention, packaging cell lines are produced using the expression cassettes of the present invention. For example, a cell line useful for packaging lentivirus vectors comprises suitable host cells that have an expression vector containing an expression cassette of the present invention wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell. In a preferred embodiment, such host cells may be transfected with one or more expression cassettes having a polynucleotide sequence that encodes an HIV polymerase polypeptide or polypeptides derived therefrom, for example, where the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6. Further, the HIV polymerase polypeptide may be modified by deletions of coding regions corresponding to reverse transcriptase and integrase. Such a polynucleotide sequence may preserve T-helper cell and CTL epitopes, for example when used in a vaccine application. In addition, the polynucleotide sequence may also include other polypeptides. Further, polynucleotide sequences encoding additional polypeptides whose expression are useful for packaging cell line function may also be utilized.

In another aspect, the present invention includes a gene delivery or vaccine vector for use in a subject, where the vector is a suitable gene delivery vector for use in the subject, and the vector comprises one or more of any of the expression cassettes of the present

invention where the polynucleotide sequences of interest are operably linked to control elements compatible with expression in the subject. Such gene delivery vectors can be used in a method of DNA immunization of a subject, 5 for example, by introducing a gene delivery vector into the subject under conditions that are compatible with expression of the expression cassette in the subject. Gene delivery vectors useful in the practice of the present invention include, but are not limited to, 10 nonviral vectors, bacterial plasmid vectors, viral vectors, particulate carriers (where the vector is coated on a polylactide co-glycolide particles, gold or tungsten particle, for example, the coated particle can be delivered to a subject cell using a gene gun), liposome 15 preparations, and viral vectors (e.g., vectors derived from alphaviruses, pox viruses, and vaccinia viruses, as well as, retroviral vectors, including, but not limited to, lentiviral vectors). Alphavirus-derived vectors include, for example, an alphavirus cDNA construct, a 20 recombinant alphavirus particle preparation and a eukaryotic layered vector initiation system. In one embodiment, the subject is a vertebrate, preferably a mammal, and in a further embodiment the subject is a human.

25 The invention further includes a method of generating an immune response in a subject, where cells of a subject are transfected with any of the above-described gene delivery vectors (e.g., alphavirus constructs; alphavirus cDNA constructs; eukaryotic 30 layered vector initiation systems (see, e.g., U.S. Patent Number 5,814,482 for description of suitable eukaryotic layered vector initiation systems); alphavirus particle

preparations; etc.) under conditions that permit the expression of a selected polynucleotide and production of a polypeptide of interest (i.e., encoded by any expression cassette of the present invention), thereby  
5 eliciting an immunological response to the polypeptide. Transfection of the cells may be performed *ex vivo* and the transfected cells are reintroduced into the subject. Alternately, or in addition, the cells may be transfected *in vivo* in the subject. The immune response may be  
10 humoral and/or cell-mediated (cellular).

Further embodiments of the present invention include purified polynucleotides. In one embodiment, the purified polynucleotide comprises a polynucleotide sequence having at least 90% sequence identity to the  
15 sequence presented as SEQ ID NO:20, and complements thereof. In another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence having at  
20 least 90% sequence identity to the sequence presented as SEQ ID NO:20, and complements thereof. In still another embodiment, the purified polynucleotide comprises a polynucleotide sequence encoding an HIV Gag polypeptide, wherein the polynucleotide sequence comprises a sequence  
25 having at least 90% sequence identity to the sequence presented as SEQ ID NO:9, and complements thereof. In further embodiments the polynucleotide sequence comprises a sequence having at least 90% sequence identity to one of the following sequences: SEQ ID NO:4, SEQ ID NO:5, SEQ  
30 ID NO:6, SEQ ID NO:7, and complements thereof.

The polynucleotides of the present invention can be produced by recombinant techniques, synthetic techniques, or combinations thereof.

These and other embodiments of the present invention will readily occur to those of ordinary skill in the art in view of the disclosure herein.

5    **BRIEF DESCRIPTION OF THE FIGURES**

Figure 1 shows the locations of the inactivation sites for the native HIV-1SF2 Gag protein coding sequence.

10    Figure 2 shows the locations of the inactivation sites for the native HIV-1SF2 Gag-protease protein coding sequence.

Figures 3A and 3B show electron micrographs of virus-like particles. Figure 3A shows immature p55Gag virus-like particles in COS-7 cells transfected with a synthetic HIV-1<sub>SF2</sub> gag construct while Figure 3B shows 15    mature (arrows) and immature VLP in cells transfected with a modified HIV-1<sub>SF2</sub> gagprotease construct (GP2, SEQ ID NO:70). Transfected cells were fixed at 24 h (gag) or 48 h (gagprotease) post-transfection and subsequently 20    analyzed by electron microscopy (magnification at 100,000X). Cells transfected with vector alone (pCMVKm2) served as negative control (data not shown).

Figure 4 presents an image of samples from a series of fractions which were electrophoresed on an 8-16% SDS 25    polyacrylamide gel and the resulting bands visualized by commassie blue staining. The results show that the native p55 Gag virus-like particles (VLPs) banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml.

30    Figure 5 presents an image similar to Figure 4 where the analysis was performed using Gag VLPs produced by a synthetic Gag expression cassette.

Figure 6 presents a comparison of the total amount of purified HIV p55 Gag from several preparations obtained from two baculovirus expression cassettes encoding native and modified Gag.

5        Figure 7 presents an alignment of modified coding sequences of the present invention including a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NO:5), and a  
10        synthetic Gag-polymerase expression cassette (SEQ ID NO:6). A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

Figure 8 presents an image of wild-type Gag-HCV core expression samples from a series of fractions which were electrophoresed on an 8-16% SDS polyacrylamide gel and  
15        the resulting bands visualized by commassie staining.

Figure 9 shows the results of Western blot analysis of the gel shown presented in Figure 8.

Figure 10 presents results similar to those shown in Figure 9. The results in Figure 10 indicate that the  
20        main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kD, which is in accordance with the predicted molecular weight of the Gag-HCV core chimeric protein.

Figures 11A to 11D present a comparison of AT  
25        content, in percent, of cDNAs corresponding to an unstable human mRNA (human IFN $\gamma$  mRNA; 11A), wild-type HIV Gag native RNA (11B), a stable human mRNA (human GAPDH mRNA; 11C), and synthetic HIV Gag RNA (11D).

Figure 12 shows the location of the inactivation  
30        sites for the native HIV-1SF2 Gag-polymerase sequence.

Figure 13A presents a vector map of pESN2dhfr.

Figure 13B presents a map of the pCMVIII vector.

Figure 14 presents a vector map of pCMV-LINK.

Figure 15 presents a schematic diagram showing the relationships between the following forms of the HIV Env polypeptide: gp160, gp140, gp120, and gp41.

5 Figure 16 depicts the nucleotide sequence of wild-type gp120 from SF162 (SEQ ID NO:30).

Figure 17 depicts the nucleotide sequence of the wild-type gp140 from SF162 (SEQ ID NO:31).

Figure 18 depicts the nucleotide sequence of the wild-type gp160 from SF162 (SEQ ID NO:32).

10 Figure 19 depicts the nucleotide sequence of the construct designated gp120.modSF162 (SEQ ID NO:33).

Figure 20 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV2 (SEQ ID NO:34).

15 Figure 21 depicts the nucleotide sequence of the construct designated gp120.modSF162.delV1/V2 (SEQ ID NO:35).

Figures 22A-H show the percent A-T content over the length of the sequences for IFN $\gamma$  (Figures 2C and 2G); native gp160 Env US4 and SF162 (Figures 2A and 2E, respectively); GAPDH (Figures 2D and 2H); and the synthetic gp160 Env for US4 and SF162 (Figures 2B and 2F, respectively).

Figure 23 depicts the nucleotide sequence of the construct designated gp140.modSF162 (SEQ ID NO:36).

25 Figure 24 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV2 (SEQ ID NO:37).

Figure 25 depicts the nucleotide sequence of the construct designated gp140.modSF162.delV1/V2 (SEQ ID NO:38).

30 Figure 26 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162 (SEQ ID NO:39).

Figure 27 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV2 (SEQ ID NO:40).

Figure 28 depicts the nucleotide sequence of the construct designated gp140.mut.modSF162.delV1/V2 (SEQ ID NO:41).

5 Figure 29 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162 (SEQ ID NO:42).

Figure 30 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV2 (SEQ ID NO:43).

10 Figure 31 depicts the nucleotide sequence of the construct designated gp140.mut7.modSF162.delV1/V2 (SEQ ID NO:44).

Figure 32 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162 (SEQ ID NO:45).

15 Figure 33 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV2 (SEQ ID NO:46).

Figure 34 depicts the nucleotide sequence of the construct designated gp140.mut8.modSF162.delV1/V2 (SEQ ID NO:47).

20 Figure 35 depicts the nucleotide sequence of the construct designated gp160.modSF162 (SEQ ID NO:48).

Figure 36 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV2 (SEQ ID NO:49).

25 Figure 37 depicts the nucleotide sequence of the construct designated gp160.modSF162.delV1/V2 (SEQ ID NO:50).

Figure 38 depicts the nucleotide sequence of the wild-type gp120 from US4 (SEQ ID NO:51).

30 Figure 39 depicts the nucleotide sequence of the wild-type gp140 from US4 (SEQ ID NO:52).

Figure 40 depicts the nucleotide sequence of the wild-type gp160 from US4 (SEQ ID NO:53).

Figure 41 depicts the nucleotide sequence of the construct designated gp120.modUS4 (SEQ ID NO:54).



Figure 42 depicts the nucleotide sequence of the construct designated gp120.modUS4.del 128-194 (SEQ ID NO:55).

5 Figure 43 depicts the nucleotide sequence of the construct designated gp140.modUS4 (SEQ ID NO:56).

Figure 44 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4 (SEQ ID NO:57).

Figure 45 depicts the nucleotide sequence of the construct designated gp140.TM.modUS4 (SEQ ID NO:58).

10 Figure 46 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV1/V2 (SEQ ID NO:59).

Figure 47 depicts the nucleotide sequence of the construct designated gp140.modUS4.delV2 (SEQ ID NO:60).

15 Figure 48 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.delV1/V2 (SEQ ID NO:61).

Figure 49 depicts the nucleotide sequence of the construct designated gp140.modUS4.del 128-194 (SEQ ID NO:62).

20 Figure 50 depicts the nucleotide sequence of the construct designated gp140.mut.modUS4.del 128-194 (SEQ ID NO:63).

Figure 51 depicts the nucleotide sequence of the construct designated gp160.modUS4 (SEQ ID NO:64).

25 Figure 52 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1 (SEQ ID NO:65).

Figure 53 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV2 (SEQ ID NO:66).

30 Figure 54 depicts the nucleotide sequence of the construct designated gp160.modUS4.delV1/V2 (SEQ ID NO:67).

Figure 55 depicts the nucleotide sequence of the construct designated gp160.modUS4.del 128-194 (SEQ ID NO:68).

5 Figure 56 depicts the nucleotide sequence of the common region of Env from wild-type US4 (SEQ ID NO:69).

Figure 57 depicts the nucleotide sequence of the common region of Env from wild-type SF162 (SEQ ID NO:70).

10 Figure 58 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from US4 (SEQ ID NO:71).

Figure 59 depicts the nucleotide sequence of synthetic sequences corresponding to the common region of Env from SF162 (SEQ ID NO:72).

15 Figure 60 presents a schematic representation of an Env polypeptide purification strategy.

Figure 61 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.Gag.modSF2 (SEQ ID NO:73).

20 Figure 62 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.Gag.modSF2 (SEQ ID NO:74).

Figure 63 depicts the nucleotide sequence of the bicistronic construct designated gp160.modUS4.-delV1/V2.Gag.modSF2 (SEQ ID NO:75).

25 Figure 64 depicts the nucleotide sequence of the bicistronic construct designated gp160.modSF162.delV2.Gag.modSF2 (SEQ ID NO:76).

30 Figures 65A-65F show micrographs of 293T cells transfected with the following polypeptide encoding sequences: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C, gp160.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and

gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and gag.modSF2.

Figures 66A and 66B present alignments of selected modified coding sequences of the present invention including a common region defined for each group of synthetic Env expression cassettes. Figure 66A presents alignments of modified SF162 sequences. Figure 66B presents alignments of modified US4 sequences. The SEQ ID NOs for these sequences are presented in Tables 1A and 1B.

Figure 67 shows the ELISA titers (binding antibodies) obtained in two rhesus macaques (H445, lines with solid black dots; and J408, lines with open squares). The y-axis is the end-point gp140 ELISA titers and the x-axis shows weeks post-immunization. The dashed lines at 0, 4, and 8 weeks represent DNA immunizations. The alternating dash/dotted line at 27 weeks indicates a DNA plus protein boost immunization.

Figure 68 (SEQ ID NO:77) depicts the wild-type nucleotide sequence of Gag reverse transcriptase from SF2.

Figure 69 (SEQ ID NO:78) depicts the nucleotide sequence of the construct designated GP1.

Figure 70 (SEQ ID NO:79) depicts the nucleotide sequence of the construct designated GP2.

Figure 71 (SEQ ID NO:80) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YM. FS(+) indicates that there is a frameshift in the GagPol coding sequence.

Figure 72 (SEQ ID NO:81) depicts the nucleotide sequence of the construct designated FS(+).protinact.RTopt.YMWM.

Figure 73 (SEQ ID NO:82) depicts the nucleotide sequence of the construct designated FS(-

).protmod.RTopt.YM. FS(-) indicates that there is no frameshift in the GagPol coding sequence.

Figure 74 (SEQ ID NO:83) depicts the nucleotide sequence of the construct designated

5 FS(-).protmod.RTopt.YMWM.

Figure 75 (SEQ ID NO:84) depicts the nucleotide sequence of the construct designated FS(-).protmod.RTopt(+).

10 Figure 76 (SEQ ID NO:85) depicts the nucleotide sequence of wild type Tat from isolate SF162.

Figure 77 (SEQ ID NO:86) depicts the amino acid sequence of the tat polypeptide.

15 Figure 78 (SEQ ID NO:87) depicts the nucleotide sequence of a synthetic Tat construct designated Tat.SF162.opt.

20 Figure 79 (SEQ ID NO:88) depicts the nucleotide sequence of a synthetic Tat construct designated tat.cys22.sf162.opt. The construct encodes a tat polypeptide in which the cystein residue at position 22 of the wild type Tat polypeptide is replaced by a glycine residue.

Figures 80A to 80E are an alignment of the nucleotide sequences of the constructs designated Gag.mod.SF2, GP1 (SEQ ID NO:78), and GP2 (SEQ ID NO:79).

25 Figure 81 (SEQ ID NO:89) depicts the nucleotide sequence of the construct designated tataminoSF162.opt, which encodes the amino terminus of that tat protein. The codon encoding the cystein-22 residue is underlined.

30 Figure 82 (SEQ ID NO:90) depicts the amino acid sequence of the polypeptide encoded by the construct designated tat.cys22.SF162.opt (SEQ ID NO:88).

## DETAILED DESCRIPTION OF THE INVENTION

The practice of the present invention will employ, unless otherwise indicated, conventional methods of chemistry, biochemistry, molecular biology, immunology and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); *Methods In Enzymology* (S. Colowick and N. Kaplan, eds., Academic Press, Inc.); and *Handbook of Experimental Immunology*, Vols. I-IV (D.M. Weir and C.C. Blackwell, eds., 1986, Blackwell Scientific Publications); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Short Protocols in Molecular Biology*, 4th ed. (Ausubel et al. eds., 1999, John Wiley & Sons); *Molecular Biology Techniques: An Intensive Laboratory Course*, (Ream et al., eds., 1998, Academic Press); *PCR (Introduction to Biotechniques Series)*, 2nd ed. (Newton & Graham eds., 1997, Springer Verlag).

As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "an antigen" includes a mixture of two or more such agents.

## 1. DEFINITIONS

In describing the present invention, the following terms will be employed, and are intended to be defined as indicated below.

"Synthetic" sequences, as used herein, refers to Env-, tat- or Gag-encoding polynucleotides whose expression has been optimized as described herein, for example, by codon substitution, deletions, replacements and/or inactivation of inhibitory sequences. "Wild-type"

or "native" sequences, as used herein, refers to polypeptide encoding sequences that are essentially as they are found in nature, e.g., Gag encoding sequences as found in the isolate HIV-1SF2 or Env encoding sequences as found in the isolates HIV-1SF162 or HIV1US4.

As used herein, the term "virus-like particle" or "VLP" refers to a nonreplicating, viral shell, derived from any of several viruses discussed further below. VLPs are generally composed of one or more viral proteins, such as, but not limited to those proteins referred to as capsid, coat, shell, surface and/or envelope proteins, or particle-forming polypeptides derived from these proteins. VLPs can form spontaneously upon recombinant expression of the protein in an appropriate expression system. Methods for producing particular VLPs are known in the art and discussed more fully below. The presence of VLPs following recombinant expression of viral proteins can be detected using conventional techniques known in the art, such as by electron microscopy, biophysical characterization, and the like. See, e.g., Baker et al., *Biophys. J.* (1991) 60:1445-1456; Hagensee et al., *J. Virol.* (1994) 68:4503-4505. For example, VLPs can be isolated by density gradient centrifugation and/or identified by characteristic density banding (e.g., Example 7). Alternatively, cryoelectron microscopy can be performed on vitrified aqueous samples of the VLP preparation in question, and images recorded under appropriate exposure conditions.

By "particle-forming polypeptide" derived from a particular viral protein is meant a full-length or near full-length viral protein, as well as a fragment thereof, or a viral protein with internal deletions, which has the ability to form VLPs under conditions that favor VLP

formation. Accordingly, the polypeptide may comprise the full-length sequence, fragments, truncated and partial sequences, as well as analogs and precursor forms of the reference molecule. The term therefore intends

5 deletions, additions and substitutions to the sequence, so long as the polypeptide retains the ability to form a VLP. Thus, the term includes natural variations of the specified polypeptide since variations in coat proteins often occur between viral isolates. The term also

10 includes deletions, additions and substitutions that do not naturally occur in the reference protein, so long as the protein retains the ability to form a VLP. Preferred substitutions are those which are conservative in nature, i.e., those substitutions that take place within a family

15 of amino acids that are related in their side chains. Specifically, amino acids are generally divided into four families: (1) acidic -- aspartate and glutamate; (2) basic -- lysine, arginine, histidine; (3) non-polar -- alanine, valine, leucine, isoleucine, proline,

20 phenylalanine, methionine, tryptophan; and (4) uncharged polar -- glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified as aromatic amino acids.

25 An "antigen" refers to a molecule containing one or more epitopes (either linear, conformational or both) that will stimulate a host's immune system to make a humoral and/or cellular antigen-specific response. The term is used interchangeably with the term "immunogen."

30 Normally, a B-cell epitope will include at least about 5 amino acids but can be as small as 3-4 amino acids. A T-cell epitope, such as a CTL epitope, will include at least about 7-9 amino acids, and a helper T-cell epitope at least about 12-20 amino acids. Normally, an epitope

will include between about 7 and 15 amino acids, such as, 9, 10, 12 or 15 amino acids. The term "antigen" denotes both subunit antigens, (i.e., antigens which are separate and discrete from a whole organism with which the antigen is associated in nature), as well as, killed, attenuated or inactivated bacteria, viruses, fungi, parasites or other microbes. Antibodies such as anti-idiotypic antibodies, or fragments thereof, and synthetic peptide mimotopes, which can mimic an antigen or antigenic determinant, are also captured under the definition of antigen as used herein. Similarly, an oligonucleotide or polynucleotide which expresses an antigen or antigenic determinant *in vivo*, such as in gene therapy and DNA immunization applications, is also included in the definition of antigen herein.

For purposes of the present invention, antigens can be derived from any of several known viruses, bacteria, parasites and fungi, as described more fully below. The term also intends any of the various tumor antigens. Furthermore, for purposes of the present invention, an "antigen" refers to a protein which includes modifications, such as deletions, additions and substitutions (generally conservative in nature), to the native sequence, so long as the protein maintains the ability to elicit an immunological response, as defined herein. These modifications may be deliberate, as through site-directed mutagenesis, or may be accidental, such as through mutations of hosts which produce the antigens.

An "immunological response" to an antigen or composition is the development in a subject of a humoral and/or a cellular immune response to an antigen present in the composition of interest. For purposes of the present invention, a "humoral immune response" refers to



an immune response mediated by antibody molecules, while a "cellular immune response" is one mediated by T-lymphocytes and/or other white blood cells. One important aspect of cellular immunity involves an antigen-specific response by cytolytic T-cells ("CTL"s). CTLs have specificity for peptide antigens that are presented in association with proteins encoded by the major histocompatibility complex (MHC) and expressed on the surfaces of cells. CTLs help induce and promote the destruction of intracellular microbes, or the lysis of cells infected with such microbes. Another aspect of cellular immunity involves an antigen-specific response by helper T-cells. Helper T-cells act to help stimulate the function, and focus the activity of, nonspecific effector cells against cells displaying peptide antigens in association with MHC molecules on their surface. A "cellular immune response" also refers to the production of cytokines, chemokines and other such molecules produced by activated T-cells and/or other white blood cells, including those derived from CD4+ and CD8+ T-cells.

A composition or vaccine that elicits a cellular immune response may serve to sensitize a vertebrate subject by the presentation of antigen in association with MHC molecules at the cell surface. The cell-mediated immune response is directed at, or near, cells presenting antigen at their surface. In addition, antigen-specific T-lymphocytes can be generated to allow for the future protection of an immunized host.

The ability of a particular antigen to stimulate a cell-mediated immunological response may be determined by a number of assays, such as by lymphoproliferation (lymphocyte activation) assays, CTL cytotoxic cell assays, or by assaying for T-lymphocytes specific for the

antigen in a sensitized subject. Such assays are well known in the art. See, e.g., Erickson et al., *J.*

*Immunol.* (1993) 151:4189-4199; Doe et al., *Eur. J.*

*Immunol.* (1994) 24:2369-2376. Recent methods of

5 measuring cell-mediated immune response include measurement of intracellular cytokines or cytokine secretion by T-cell populations, or by measurement of epitope specific T-cells (e.g., by the tetramer technique) (reviewed by McMichael, A.J., and O'Callaghan, C.A., *J. Exp. Med.* 187(9)1367-1371, 1998; Mcheyzer-Williams, M.G., et al, *Immunol. Rev.* 150:5-21, 1996; Lalvani, A., et al, *J. Exp. Med.* 186:859-865, 1997).

Thus, an immunological response as used herein may be one which stimulates the production of CTLs, and/or  
15 the production or activation of helper T- cells. The antigen of interest may also elicit an antibody-mediated immune response. Hence, an immunological response may include one or more of the following effects: the production of antibodies by B-cells; and/or the  
20 activation of suppressor T-cells and/or  $\gamma\delta$  T-cells directed specifically to an antigen or antigens present in the composition or vaccine of interest. These responses may serve to neutralize infectivity, and/or mediate antibody-complement, or antibody dependent cell  
25 cytotoxicity (ADCC) to provide protection to an immunized host. Such responses can be determined using standard immunoassays and neutralization assays, well known in the art.

An "immunogenic composition" is a composition that  
30 comprises an antigenic molecule where administration of the composition to a subject results in the development in the subject of a humoral and/or a cellular immune response to the antigenic molecule of interest.

By "subunit vaccine" is meant a vaccine composition which includes one or more selected antigens but not all antigens, derived from or homologous to, an antigen from a pathogen of interest such as from a virus, bacterium, parasite or fungus. Such a composition is substantially free of intact pathogen cells or pathogenic particles, or the lysate of such cells or particles. Thus, a "subunit vaccine" can be prepared from at least partially purified (preferably substantially purified) immunogenic polypeptides from the pathogen, or analogs thereof. The method of obtaining an antigen included in the subunit vaccine can thus include standard purification techniques, recombinant production, or synthetic production.

"Substantially purified" generally refers to isolation of a substance (compound, polynucleotide, protein, polypeptide, polypeptide composition) such that the substance comprises the majority percent of the sample in which it resides. Typically in a sample a substantially purified component comprises 50%, preferably 80%-85%, more preferably 90-95% of the sample. Techniques for purifying polynucleotides and polypeptides of interest are well-known in the art and include, for example, ion-exchange chromatography, affinity chromatography and sedimentation according to density.

A "coding sequence" or a sequence which "encodes" a selected polypeptide, is a nucleic acid molecule which is transcribed (in the case of DNA) and translated (in the case of mRNA) into a polypeptide *in vivo* when placed under the control of appropriate regulatory sequences (or "control elements"). The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A coding sequence can include, but

is not limited to, cDNA from viral, procaryotic or eucaryotic mRNA, genomic DNA sequences from viral or procaryotic DNA, and even synthetic DNA sequences. A transcription termination sequence may be located 3' to the coding sequence.

Typical "control elements", include, but are not limited to, transcription promoters, transcription enhancer elements, transcription termination signals, polyadenylation sequences (located 3' to the translation stop codon), sequences for optimization of initiation of translation (located 5' to the coding sequence), and translation termination sequences, see e.g., McCaughan et al. (1995) *PNAS USA* 92:5431-5435; Kochetov et al (1998) *FEBS Letts.* 440:351-355.

A "nucleic acid" molecule can include, but is not limited to, procaryotic sequences, eucaryotic mRNA, cDNA from eucaryotic mRNA, genomic DNA sequences from eucaryotic (e.g., mammalian) DNA, and even synthetic DNA sequences. The term also captures sequences that include any of the known base analogs of DNA and RNA.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a coding sequence is capable of effecting the expression of the coding sequence when the proper enzymes are present. The promoter need not be contiguous with the coding sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

"Recombinant" as used herein to describe a nucleic acid molecule means a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of the polynucleotide with which it is associated in nature; and/or (2) is linked to a polynucleotide other than that to which it is linked in nature. The term "recombinant" as used with respect to a protein or polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. "Recombinant host cells," "host cells," "cells," "cell lines," "cell cultures," and other such terms denoting procaryotic microorganisms or eucaryotic cell lines cultured as unicellular entities, are used interchangeably, and refer to cells which can be, or have been, used as recipients for recombinant vectors or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell which are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired peptide, are included in the progeny intended by this definition, and are covered by the above terms.

Techniques for determining amino acid sequence "similarity" are well known in the art. In general, "similarity" means the exact amino acid to amino acid comparison of two or more polypeptides at the appropriate place, where amino acids are identical or possess similar chemical and/or physical properties such as charge or hydrophobicity. A so-termed "percent similarity" then

can be determined between the compared polypeptide sequences. Techniques for determining nucleic acid and amino acid sequence identity also are well known in the art and include determining the nucleotide sequence of the mRNA for that gene (usually via a cDNA intermediate) and determining the amino acid sequence encoded thereby, and comparing this to a second amino acid sequence. In general, "identity" refers to an exact nucleotide to nucleotide or amino acid to amino acid correspondence of two polynucleotides or polypeptide sequences, respectively.

Two or more polynucleotide sequences can be compared by determining their "percent identity." Two or more amino acid sequences likewise can be compared by determining their "percent identity." The percent identity of two sequences, whether nucleic acid or peptide sequences, is generally described as the number of exact matches between two aligned sequences divided by the length of the shorter sequence and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981). This algorithm can be extended to use with peptide sequences using the scoring matrix developed by Dayhoff, *Atlas of Protein Sequences and Structure*, M.O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986). An implementation of this algorithm for nucleic acid and peptide sequences is provided by the Genetics Computer Group (Madison, WI) in their BestFit utility application. The default parameters for this method are

described in the Wisconsin Sequence Analysis Package Program Manual, Version 8 (1995) (available from Genetics Computer Group, Madison, WI). Other equally suitable programs for calculating the percent identity or  
5 similarity between sequences are generally known in the art.

For example, percent identity of a particular nucleotide sequence to a reference sequence can be determined using the homology algorithm of Smith and  
10 Waterman with a default scoring table and a gap penalty of six nucleotide positions. Another method of establishing percent identity in the context of the present invention is to use the MPSRCH package of programs copyrighted by the University of Edinburgh,  
15 developed by John F. Collins and Shane S. Sturrok, and distributed by IntelliGenetics, Inc. (Mountain View, CA). From this suite of packages, the Smith-Waterman algorithm can be employed where default parameters are used for the scoring table (for example, gap open penalty of 12, gap  
20 extension penalty of one, and a gap of six). From the data generated, the "Match" value reflects "sequence identity." Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, such as the alignment program  
25 BLAST, which can also be used with default parameters. For example, BLASTN and BLASTP can be used with the following default parameters: genetic code = standard; filter = none; strand = both; cutoff = 60; expect = 10; Matrix = BLOSUM62; Descriptions = 50 sequences; sort by =  
30 HIGH SCORE; Databases = non-redundant, GenBank + EMBL + DDBJ + PDB + GenBank CDS translations + Swiss protein + Spupdate + PIR. Details of these programs can be found at

the following internet address:  
<http://www.ncbi.nlm.gov/cgi-bin/BLAST>.

One of skill in the art can readily determine the proper search parameters to use for a given sequence in the above programs. For example, the search parameters may vary based on the size of the sequence in question. Thus, for example, a representative embodiment of the present invention would include an isolated

polynucleotide having X contiguous nucleotides, wherein  
10 (i) the X contiguous nucleotides have at least about 50% identity to Y contiguous nucleotides derived from any of the sequences described herein, (ii) X equals Y, and (iii) X is greater than or equal to 6 nucleotides and up to 5000 nucleotides, preferably greater than or equal to  
15 8 nucleotides and up to 5000 nucleotides, more preferably 10-12 nucleotides and up to 5000 nucleotides, and even more preferably 15-20 nucleotides, up to the number of nucleotides present in the full-length sequences described herein (e.g., see the Sequence Listing and  
20 claims), including all integer values falling within the above-described ranges.

The synthetic expression cassettes (and purified polynucleotides) of the present invention include related polynucleotide sequences having about 80% to 100%,  
25 greater than 80-85%, preferably greater than 90-92%, more preferably greater than 95%, and most preferably greater than 98% sequence (including all integer values falling within these described ranges) identity to the synthetic expression cassette sequences disclosed herein (for  
30 example, to the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.



Two nucleic acid fragments are considered to "selectively hybridize" as described herein. The degree of sequence identity between two nucleic acid molecules affects the efficiency and strength of hybridization events between such molecules. A partially identical nucleic acid sequence will at least partially inhibit a completely identical sequence from hybridizing to a target molecule. Inhibition of hybridization of the completely identical sequence can be assessed using hybridization assays that are well known in the art (e.g., Southern blot, Northern blot, solution hybridization, or the like, see Sambrook, et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, (1989) Cold Spring Harbor, N.Y.). Such assays can be conducted using varying degrees of selectivity, for example, using conditions varying from low to high stringency. If conditions of low stringency are employed, the absence of non-specific binding can be assessed using a secondary probe that lacks even a partial degree of sequence identity (for example, a probe having less than about 30% sequence identity with the target molecule), such that, in the absence of non-specific binding events, the secondary probe will not hybridize to the target.

When utilizing a hybridization-based detection system, a nucleic acid probe is chosen that is complementary to a target nucleic acid sequence, and then by selection of appropriate conditions the probe and the target sequence "selectively hybridize," or bind, to each other to form a hybrid molecule. A nucleic acid molecule that is capable of hybridizing selectively to a target sequence under "moderately stringent" typically

hybridizes under conditions that allow detection of a target nucleic acid sequence of at least about 10-14 nucleotides in length having at least approximately 70% sequence identity with the sequence of the selected nucleic acid probe. Stringent hybridization conditions typically allow detection of target nucleic acid sequences of at least about 10-14 nucleotides in length having a sequence identity of greater than about 90-95% with the sequence of the selected nucleic acid probe.

Hybridization conditions useful for probe/target hybridization where the probe and target have a specific degree of sequence identity, can be determined as is known in the art (see, for example, Nucleic Acid Hybridization: A Practical Approach, editors B.D. Hames and S.J. Higgins, (1985) Oxford; Washington, DC; IRL Press).

With respect to stringency conditions for hybridization, it is well known in the art that numerous equivalent conditions can be employed to establish a particular stringency by varying, for example, the following factors: the length and nature of probe and target sequences, base composition of the various sequences, concentrations of salts and other hybridization solution components, the presence or absence of blocking agents in the hybridization solutions (e.g., formamide, dextran sulfate, and polyethylene glycol), hybridization reaction temperature and time parameters, as well as, varying wash conditions. The selection of a particular set of hybridization conditions is selected following standard methods in the art (see, for example, Sambrook, et al., Molecular Cloning: A

Laboratory Manual, Second Edition, (1989) Cold Spring Harbor, N.Y.).

5 A first polynucleotide is "derived from" second polynucleotide if it has the same or substantially the same basepair sequence as a region of the second polynucleotide, its cDNA, complements thereof, or if it displays sequence identity as described above.

10 A first polypeptide is "derived from" a second polypeptide if it is (i) encoded by a first polynucleotide derived from a second polynucleotide, or (ii) displays sequence identity to the second polypeptides as described above.

15 Generally, a viral polypeptide is "derived from" a particular polypeptide of a virus (viral polypeptide) if it is (i) encoded by an open reading frame of a polynucleotide of that virus (viral polynucleotide), or (ii) displays sequence identity to polypeptides of that virus as described above.

20 "Encoded by" refers to a nucleic acid sequence which codes for a polypeptide sequence, wherein the polypeptide sequence or a portion thereof contains an amino acid sequence of at least 3 to 5 amino acids, more preferably at least 8 to 10 amino acids, and even more preferably at least 15 to 20 amino acids from a polypeptide encoded by  
25 the nucleic acid sequence. Also encompassed are polypeptide sequences which are immunologically identifiable with a polypeptide encoded by the sequence.

30 "Purified polynucleotide" refers to a polynucleotide of interest or fragment thereof which is essentially free, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, of the protein with which the polynucleotide is naturally associated. Techniques for purifying polynucleotides of interest are well-known in the art and include, for

example, disruption of the cell containing the polynucleotide with a chaotropic agent and separation of the polynucleotide(s) and proteins by ion-exchange chromatography, affinity chromatography and sedimentation according to density.

By "nucleic acid immunization" is meant the introduction of a nucleic acid molecule encoding one or more selected antigens into a host cell, for the *in vivo* expression of an antigen, antigens, an epitope, or epitopes. The nucleic acid molecule can be introduced directly into a recipient subject, such as by injection, inhalation, oral, intranasal and mucosal administration, or the like, or can be introduced *ex vivo*, into cells which have been removed from the host. In the latter case, the transformed cells are reintroduced into the subject where an immune response can be mounted against the antigen encoded by the nucleic acid molecule.

"Gene transfer" or "gene delivery" refers to methods or systems for reliably inserting DNA or RNA of interest into a host cell. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred replicons (e.g., episomes), or integration of transferred genetic material into the genomic DNA of host cells. Gene delivery expression vectors include, but are not limited to, vectors derived from bacterial plasmid vectors, viral vectors, non-viral vectors, alphaviruses, pox viruses and vaccinia viruses. When used for immunization, such gene delivery expression vectors may be referred to as vaccines or vaccine vectors.

"T lymphocytes" or "T cells" are non-antibody producing lymphocytes that constitute a part of the cell-mediated arm of the immune system. T cells arise from immature lymphocytes that migrate from the bone marrow to

the thymus, where they undergo a maturation process under the direction of thymic hormones. Here, the mature lymphocytes rapidly divide increasing to very large numbers. The maturing T cells become immunocompetent  
5 based on their ability to recognize and bind a specific antigen. Activation of immunocompetent T cells is triggered when an antigen binds to the lymphocyte's surface receptors.

The term "transfection" is used to refer to the  
10 uptake of foreign DNA by a cell. A cell has been "transfected" when exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are generally known in the art. See, e.g., Graham et al. (1973) *Virology*, 52:456, Sambrook et al.  
15 (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, and Chu et al. (1981) *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into  
20 suitable host cells. The term refers to both stable and transient uptake of the genetic material, and includes uptake of peptide- or antibody-linked DNAs.

A "vector" is capable of transferring gene sequences to target cells (e.g., bacterial plasmid vectors, viral  
25 vectors, non-viral vectors, particulate carriers, and liposomes). Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to  
30 target cells. Thus, the term includes cloning and expression vehicles, as well as viral vectors.

Transfer of a "suicide gene" (e.g., a drug-susceptibility gene) to a target cell renders the cell sensitive to compounds or compositions that are

relatively nontoxic to normal cells. Moolten, F.L. (1994) *Cancer Gene Ther.* 1:279-287. Examples of suicide genes are thymidine kinase of herpes simplex virus (HSV-tk), cytochrome P450 (Manome et al. (1996) *Gene Therapy* 3:513-520), human deoxycytidine kinase (Manome et al. (1996) *Nature Medicine* 2(5):567-573) and the bacterial enzyme cytosine deaminase (Dong et al. (1996) *Human Gene Therapy* 7:713-720). Cells which express these genes are rendered sensitive to the effects of the relatively nontoxic prodrugs ganciclovir (HSV-tk), cyclophosphamide (cytochrome P450 2B1), cytosine arabinoside (human deoxycytidine kinase) or 5-fluorocytosine (bacterial cytosine deaminase). Culver et al. (1992) *Science* 256:1550-1552, Huber et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:8302-8306.

A "selectable marker" or "reporter marker" refers to a nucleotide sequence included in a gene transfer vector that has no therapeutic activity, but rather is included to allow for simpler preparation, manufacturing, characterization or testing of the gene transfer vector.

A "specific binding agent" refers to a member of a specific binding pair of molecules wherein one of the molecules specifically binds to the second molecule through chemical and/or physical means. One example of a specific binding agent is an antibody directed against a selected antigen.

By "subject" is meant any member of the subphylum chordata, including, without limitation, humans and other primates, including non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats and guinea pigs; birds, including domestic, wild and game birds such

as chickens, turkeys and other gallinaceous birds, ducks, geese, and the like. The term does not denote a particular age. Thus, both adult and newborn individuals are intended to be covered. The system described above  
5 is intended for use in any of the above vertebrate species, since the immune systems of all of these vertebrates operate similarly.

By "pharmaceutically acceptable" or "pharmacologically acceptable" is meant a material which  
10 is not biologically or otherwise undesirable, i.e., the material may be administered to an individual in a formulation or composition without causing any undesirable biological effects or interacting in a deleterious manner with any of the components of the  
15 composition in which it is contained.

By "physiological pH" or a "pH in the physiological range" is meant a pH in the range of approximately 7.2 to 8.0 inclusive, more typically in the range of approximately 7.2 to 7.6 inclusive.

20 As used herein, "treatment" refers to any of (i) the prevention of infection or reinfection, as in a traditional vaccine, (ii) the reduction or elimination of symptoms, and (iii) the substantial or complete elimination of the pathogen in question. Treatment may  
25 be effected prophylactically (prior to infection) or therapeutically (following infection).

"Lentiviral vector", and "recombinant lentiviral vector" are derived from the subset of retroviral vectors known as lentiviruses. Lentiviral vectors refer to a  
30 nucleic acid construct which carries, and within certain embodiments, is capable of directing the expression of a nucleic acid molecule of interest. The lentiviral vector includes at least one transcriptional promoter/enhancer or locus defining element(s), or other elements which

control gene expression by other means such as alternate splicing, nuclear RNA export, post-translational modification of messenger, or post-transcriptional modification of protein. Such vector constructs must  
5 also include a packaging signal, long terminal repeats (LTRS) or portion thereof, and positive and negative strand primer binding sites appropriate to the lentiviral vector used (if these are not already present in the retroviral vector). Optionally, the recombinant  
10 lentiviral vector may also include a signal which directs polyadenylation, selectable markers such as Neo, TK, hygromycin, phleomycin, histidinol, or DHFR, as well as one or more restriction sites and a translation termination sequence. By way of example, such vectors  
15 typically include a 5' LTR, a tRNA binding site, a packaging signal, an origin of second strand DNA synthesis, and a 3'LTR or a portion thereof.

"Lentiviral vector particle" as utilized within the present invention refers to a lentivirus which carries at  
20 least one gene of interest. The retrovirus may also contain a selectable marker. The recombinant lentivirus is capable of reverse transcribing its genetic material (RNA) into DNA and incorporating this genetic material into a host cell's DNA upon infection. Lentiviral vector  
25 particles may have a lentiviral envelope, a non-lentiviral envelope (e.g., an amphi or VSV-G envelope), or a chimeric envelope.

"Nucleic acid expression vector" or "Expression cassette" refers to an assembly which is capable of  
30 directing the expression of a sequence or gene of interest. The nucleic acid expression vector includes a promoter which is operably linked to the sequences or gene(s) of interest. Other control elements may be present as well. Expression cassettes described herein



may be contained within a plasmid construct. In addition to the components of the expression cassette, the plasmid construct may also include a bacterial origin of replication, one or more selectable markers, a signal  
5 which allows the plasmid construct to exist as single-stranded DNA (e.g., a M13 origin of replication), a multiple cloning site, and a "mammalian" origin of replication (e.g., a SV40 or adenovirus origin of replication).

10 "Packaging cell" refers to a cell which contains those elements necessary for production of infectious recombinant retrovirus (e.g., lentivirus) which are lacking in a recombinant retroviral vector. Typically, such packaging cells contain one or more expression  
15 cassettes which are capable of expressing proteins which encode *Gag*, *pol* and *env* proteins.

"Producer cell" or "vector producing cell" refers to a cell which contains all elements necessary for production of recombinant retroviral vector particles.

20

## 2. MODES OF CARRYING OUT THE INVENTION

Before describing the present invention in detail, it is to be understood that this invention is not limited to particular formulations or process parameters as such  
25 may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments of the invention only, and is not intended to be limiting.

Although a number of methods and materials similar  
30 or equivalent to those described herein can be used in the practice of the present invention, the preferred materials and methods are described herein.

## 2.1 SYNTHETIC EXPRESSION CASSETTES

### 2.1.1 MODIFICATION OF HIV-1 GAG NUCLEIC ACID CODING SEQUENCES

One aspect of the present invention is the generation of HIV-1 Gag protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. An exemplary embodiment of the present invention is illustrated herein modifying the Gag protein wild-type sequences obtained from the HIV-1SF2 strain (SEQ ID NO:1; Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November 18, 1997). Gag sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include, but are not limited to, Gag protein encoding sequences obtained from the isolates HIV<sub>111b</sub>, HIV<sub>SF2</sub>, HIV-1<sub>SF162</sub>, HIV-1<sub>SF170</sub>, HIV<sub>LAV</sub>, HIV<sub>LA1</sub>, HIV<sub>MN</sub>, HIV-1<sub>CM235</sub>, HIV-1<sub>US4</sub>, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2<sub>UC1</sub> and HIV-2<sub>UC2</sub>), and simian immunodeficiency virus (SIV). (See, e.g., *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991); *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA; for a description of these and other related viruses).

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet.

The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag coding sequences were modified to be comparable to codon usage found in highly expressed human genes. In Figure 11 (Example 1), the percent A-T content of cDNA sequences corresponding to the mRNA for a known unstable mRNA and a known stable mRNA are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA sequence of the present invention. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) relative to the native Gag sequences. The data in Figure 11 suggest that one reason for this increased production is increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag coding sequences (Example 1). The RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating effects of the INS. To overcome the post-transcriptional activating mechanisms of RRE and Rev, the instability elements were inactivated by introducing multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects. The resulting modified coding sequences are

presented as a synthetic Gag expression cassette (SEQ ID NO:4).

Modification of the Gag polypeptide coding sequences resulted in improved expression relative to the wild-type coding sequences in a number of mammalian cell lines (as well as other types of cell lines, including, but not limited to, insect cells). Further, expression of the sequences resulted in production of virus-like particles (VLPs) by these cell lines (see below). Similar Gag polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, strains, etc.) including, but not limited to such other variants include, but are not limited to, Gag polypeptide encoding sequences obtained from the isolates HIV<sub>IIIB</sub>, HIV<sub>SF2</sub>, HIV-1<sub>SF162</sub>, HIV-1<sub>SF170</sub>, HIV<sub>LAV</sub>, HIV<sub>LAI</sub>, HIV<sub>MN</sub>, HIV-1<sub>CM235</sub>, HIV-1<sub>USA</sub>, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2<sub>UC1</sub> and HIV-2<sub>UC2</sub>), and simian immunodeficiency virus (SIV). (See, e.g., Virology, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA). Gag polypeptide encoding sequences derived from these variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 1).

#### 2.1.2 FURTHER MODIFICATION OF SEQUENCES INCLUDING HIV-1 GAG NUCLEIC ACID CODING SEQUENCES

Experiments performed in support of the present invention have shown that similar modifications of HIV-1 Gag-protease, Gag-reverse transcriptase and Gag-polymerase sequences also result in improved expression

of the polyproteins, as well as, the production of VLPs formed by polypeptides produced from such modified coding sequences.

For the Gag-protease sequence (wild type, SEQ ID NO:2; modified, SEQ ID NOs:5, 78, 79), the changes in codon usage were restricted to the regions upstream of the -1 frameshift (Figure 2). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). Exemplary constructs (which include the -1 frameshift) encoding modified Gag-protease sequences include those shown in SEQ ID NOs:78 and 79 (Figures 69 and 70). These are: GP1 (SEQ ID NO:78) in which the protease region was also codon optimized and INS inactivated and GP2 (SEQ ID NO:79), in which the protease region was only subjected to INS inactivation.

For other Gag-containing sequences, for example the Gag-polymerase sequence (wild type, SEQ ID NO:3; modified, SEQ ID NO:6) or Gag-reverse transcriptase (wild type, SEQ ID NO:77; modified SEQ ID NOs:80-84), the changes in codon usage are similar to those for the Gag-protease sequence. Those expression cassettes which contain a frameshift in the GagPol coding sequence are designated "FS(+)" (SEQ ID NOs:80 and 81, Figures 71 and 72) while the designation "FS(-)" (SEQ ID Nos: 82, 83 and 84, Figures 73, 74 and 75) indicates that there is no frameshift utilized in this coding sequence.

In addition to polyproteins containing HIV-related sequences, the various Gag-, Gag-prot, Gag-pol, Gag-reverse transcriptase encoding sequences of the present invention can be fused to other polypeptides (creating chimeric polypeptides) for which an immunogenic response is desired. An example of such a chimeric protein is the

joining of the improved expression Gag encoding sequences to the Hepatitis C Virus (HCV) core protein. In this case, the HCV-core encoding sequences were placed in-frame with the HIV-Gag encoding sequences, resulting in the Gag/HCV-core encoding sequence presented as SEQ ID NO:7 (wild type sequence presented as SEQ ID NO:8).

Further sequences useful in the practice of the present invention include, but are not limited to, sequences encoding viral epitopes/antigens (including but not limited to, HCV antigens (e.g., E1, E2; Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997), HIV antigens (e.g., derived from *nef*, *tat*, *rev*, *vpu*, *vif*, *vpr* and/or *env*); and sequences encoding tumor antigens/epitopes. Additional sequences are described below. Also, variations on the orientation of the Gag and other coding sequences, relative to each other, are also described below.

Gag, Gag-protease, Gag-reverse transcriptase and/or Gag-polymerase polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV<sub>IIIB</sub>, HIV<sub>SF2</sub>, HIV<sub>SF162</sub>, HIV<sub>US4</sub>, HIV<sub>CM235</sub>, HIV<sub>LAV</sub>, HIV<sub>LA1</sub>, HIV<sub>MN</sub> (see, e.g., Myers et al. Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic expression cassettes can be generated using

such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes of the present invention include related Gag polypeptide coding sequences having greater than 75%, preferably greater than 80-85%, more preferably greater than 90-95%, and most preferably greater than 98% sequence identity (or any integer value within these ranges) to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NO:4; SEQ ID NO:5; SEQ ID NO:6; and SEQ ID NO:20, the Gag Major Homology Region).

#### 2.1.3 EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1 GAG AND RELATED POLYPEPTIDES

Several synthetic Gag-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to evaluate levels of expression and production of VLPs. Two modified synthetic coding sequences are presented as a synthetic Gag expression cassette (SEQ ID NO:4) and a synthetic Gag-protease expression cassette (SEQ ID NOs:78 and 79). Other synthetic Gag-encoding proteins are presented, for example, as SEQ ID NOs:80 through 84. The synthetic DNA fragments for Gag-encoding polypeptides (e.g., Gag, Gag-protease, Gag-polymerase, Gag-reverse transcriptase) were cloned into expression vectors described in Example 1, including, a transient expression vector, CMV-promoter-based mammalian vectors, and a shuttle vector for use in baculovirus expression systems. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several different cell types, including a variety of mammalian

cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of p24 (Gag) expression in supernatants were  
5 evaluated (Example 2). The results of these assays demonstrated that expression of synthetic Gag-encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Table 2).

Further, Western Blot analysis showed that cells  
10 containing the synthetic Gag expression cassette produced the expected 55 kD (p55) protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of  
15 production were significantly higher in cell supernatants for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassettes  
20 produced the expected Gag-prot protein at comparably higher per-cell concentrations than cells containing the wild-type expression cassette.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Gag expression  
25 cassette showed that it provides superior production of both p55 protein and VLPs, relative to the wild-type Gag sequences (Examples 6 and 7).

Efficient expression of these Gag-containing polypeptides in mammalian cell lines provides the  
30 following benefits: the Gag polypeptides are free of baculovirus contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Gag-containing polypeptides in



CHO or other mammalian cells which is not feasible in the absence of the increased expression obtained using the constructs of the present invention. Exemplary Mammalian cell lines include, but are not limited to, BHK, VERO, HT1080, 293, 293T, RD, COS-7, CHO, Jurkat, HUT, SUPT, C8166, MOLT4/clone8, MT-2, MT-4, H9, PM1, CEM, myeloma cells (e.g., SB20 cells) and CEMX174, such cell lines are available, for example, from the A.T.C.C.).

A synthetic Gag expression cassette of the present invention also demonstrated high levels of expression and VLP production when transfected into insect cells (Example 7). Further, in addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of contaminating baculovirus proteins than the final purified product from the native p55-expressed Gag.

Further, synthetic Gag expression cassettes of the present invention have also been introduced into yeast vectors which were transformed into and efficiently expressed by yeast cells (*Saccharomyces cerevisiae*; using vectors as described in Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998).

In addition to the mammalian and insect vectors described in the Examples, the synthetic expression cassettes of the present invention can be incorporated into a variety of expression vectors using selected expression control elements. Appropriate vectors and control elements for any given cell type can be selected by one having ordinary skill in the art in view of the teachings of the present specification and information known in the art about expression vectors.

For example, a synthetic Gag expression cassette can be inserted into a vector which includes control elements operably linked to the desired coding sequence, which

allow for the expression of the gene in a selected cell-type. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter (a CMV promoter can include intron A), RSV, HIV-LTR, the mouse mammary tumor virus LTR promoter (MMLV-LTR), FIV-LTR, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook, et al., *supra*, as well as a bovine growth hormone terminator sequence. Introns, containing splice donor and acceptor sites, may also be designed into the constructs for use with the present invention (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986).

The desired synthetic Gag polypeptide encoding sequences can be cloned into any number of commercially available vectors to generate expression of the polypeptide in an appropriate host system. These systems include, but are not limited to, the following:

5 baculovirus expression {Reilly, P.R., et al., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992); Beames, et al., *Biotechniques* 11:378 (1991); Pharmingen; Clontech, Palo Alto, CA}}, vaccinia expression {Earl, P. L., et al.,

10 "Expression of proteins in mammalian cells using vaccinia" In *Current Protocols in Molecular Biology* (F. M. Ausubel, et al. Eds.), Greene Publishing Associates & Wiley Interscience, New York (1991); Moss, B., et al., U.S. Patent Number 5,135,855, issued 4 August 1992},

15 expression in bacteria {Ausubel, F.M., et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley and Sons, Inc., Media PA; Clontech}, expression in yeast {Rosenberg, S. and Tekamp-Olson, P., U.S. Patent No. RE35,749, issued, March 17, 1998; Shuster, J.R., U.S. Patent No. 5,629,203,

20 issued May 13, 1997; Gellissen, G., et al., *Antonie Van Leeuwenhoek*, 62(1-2):79-93 (1992); Romanos, M.A., et al., *Yeast* 8(6):423-488 (1992); Goeddel, D.V., *Methods in Enzymology* 185 (1990); Guthrie, C., and G.R. Fink, *Methods in Enzymology* 194 (1991)}, expression in

25 mammalian cells {Clontech; Gibco-BRL, Ground Island, NY; e.g., Chinese hamster ovary (CHO) cell lines (Haynes, J., et al., *Nuc. Acid. Res.* 11:687-706 (1983); 1983, Lau, Y.F., et al., *Mol. Cell. Biol.* 4:1469-1475 (1984); Kaufman, R. J., "Selection and coamplification of

30 heterologous genes in mammalian cells," in *Methods in Enzymology*, vol. 185, pp537-566. Academic Press, Inc., San Diego CA (1991)}, and expression in plant cells {plant cloning vectors, Clontech Laboratories, Inc., Palo Alto, CA, and Pharmacia LKB Biotechnology, Inc.,

Piscataway, NJ; Hood, E., et al., *J. Bacteriol.* 168:1291-1301 (1986); Nagel, R., et al., *FEMS Microbiol. Lett.* 67:325 (1990); An, et al., "Binary Vectors", and others in Plant Molecular Biology Manual A3:1-19 (1988);  
5 Miki, B.L.A., et al., pp.249-265, and others in Plant DNA Infectious Agents (Hohn, T., et al., eds.) Springer-Verlag, Wien, Austria, (1987); *Plant Molecular Biology: Essential Techniques*, P.G. Jones and J.M. Sutton, New York, J. Wiley, 1997; Miglani, Gurbachan *Dictionary of*  
10 *Plant Genetics and Molecular Biology*, New York, Food Products Press, 1998; Henry, R. J., *Practical Applications of Plant Molecular Biology*, New York, Chapman & Hall, 1997}.

Also included in the invention is an expression  
15 vector, such as the CMV promoter-containing vectors described in Example 1, containing coding sequences and expression control elements which allow expression of the coding regions in a suitable host. The control elements generally include a promoter, translation initiation  
20 codon, and translation and transcription termination sequences, and an insertion site for introducing the insert into the vector. Translational control elements have been reviewed by M. Kozak (e.g., Kozak, M., *Mamm. Genome* 7(8):563-574, 1996; Kozak, M., *Biochimie*  
25 76(9):815-821, 1994; Kozak, M., *J Cell Biol* 108(2):229-241, 1989; Kozak, M., and Shatkin, A.J., *Methods Enzymol* 60:360-375, 1979).

Expression in yeast systems has the advantage of commercial production. Recombinant protein production by  
30 vaccinia and CHO cell line have the advantage of being mammalian expression systems. Further, vaccinia virus expression has several advantages including the following: (i) its wide host range; (ii) faithful post-

transcriptional modification, processing, folding, transport, secretion, and assembly of recombinant proteins; (iii) high level expression of relatively soluble recombinant proteins; and (iv) a large capacity to accommodate foreign DNA.

The recombinantly expressed polypeptides from synthetic Gag-encoding expression cassettes are typically isolated from lysed cells or culture media. Purification can be carried out by methods known in the art including salt fractionation, ion exchange chromatography, gel filtration, size-exclusion chromatography, size-fractionation, and affinity chromatography.

Immunoaffinity chromatography can be employed using antibodies generated based on, for example, Gag antigens.

Advantages of expressing the Gag-containing proteins of the present invention using mammalian cells include, but are not limited to, the following: well-established protocols for scale-up production; the ability to produce VLPs; cell lines are suitable to meet good manufacturing process (GMP) standards; culture conditions for mammalian cells are known in the art.

#### 2.1.4 MODIFICATION OF HIV-1 ENV NUCLEIC ACID CODING SEQUENCES

One aspect of the present invention is the generation of HIV-1 Env protein coding sequences, and related sequences, having improved expression relative to the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the Env protein wild-type sequences obtained from the HIV-1 subtype B strains HIV-1US4 and HIV-1SF162 (Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos,

New Mexico: Los Alamos National Laboratory). Env sequence obtained from other HIV variants may be manipulated in similar fashion following the teachings of the present specification. Such other variants include those  
5 described above in Section 2.1.1 and on the World Wide Web (Internet), for example at <http://hiv-web.lanl.gov/cgi-bin/hivDB3/public/wdb/ssampublic> and <http://hiv-web.lanl.gov>.

First, the HIV-1 codon usage pattern was modified so  
10 that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes (Example 1). The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content  
15 in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly  
20 expressed human genes. Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) relative to the native Env sequences. One reason for this increased production may  
25 be increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

Modification of the Env polypeptide coding sequences resulted in improved expression relative to the wild-type  
30 coding sequences in a number of mammalian cell lines. Similar Env polypeptide coding sequences can be obtained from a variety of isolates (families, sub-types, etc.). Env polypeptide encoding sequences derived from these variants can be optimized and tested for improved

expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

5           2.1.5       FURTHER MODIFICATION OF HIV-1 ENV NUCLEIC ACID  
                  CODING SEQUENCES

          In addition to proteins containing HIV-related sequences, the Env encoding sequences of the present invention can be fused to other polypeptides (creating  
10   chimeric polypeptides). Also, variations on the orientation of the Env and other coding sequences, relative to each other, are contemplated. Further, the HIV protein encoding cassettes of the present invention can be co-expressed using one vector or multiple vectors.  
15   In addition, the polyproteins can be operably linked to the same or different promoters.

          Env polypeptide coding sequences can be obtained from any HIV isolates (different families, subtypes, and strains) including but not limited to the isolates HIV<sub>IIIB</sub>,  
20   HIV<sub>SF2</sub>, HIV<sub>US4</sub>, HIV<sub>CM235</sub>, HIV<sub>SF162</sub>, HIV<sub>LAV</sub>, HIV<sub>LA1</sub>, HIV<sub>MN</sub>) (see, e.g., Myers et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico (1992); Myers et al., *Human Retroviruses and Aids*, 1997, Los Alamos, New Mexico: Los Alamos National Laboratory). Synthetic  
25   expression cassettes can be generated using such coding sequences as starting material by following the teachings of the present specification (e.g., see Example 1). Further, the synthetic expression cassettes (and purified polynucleotides) of the present invention include related  
30   Env polypeptide coding sequences having greater than 90%, preferably greater than 92%, more preferably greater than 95%, and most preferably greater than 98% sequence identity to the synthetic expression cassette sequences disclosed herein (for example, SEQ ID NOs:71-72; and/or

the sequences presented in Tables 1A and 1B) when the sequences of the present invention are used as the query sequence.

5           2.1.6           EXPRESSION OF SYNTHETIC SEQUENCES ENCODING HIV-1  
                          ENV AND RELATED POLYPEPTIDES

Several synthetic Env-encoding sequences (expression cassettes) of the present invention were cloned into a number of different expression vectors (Example 1) to  
10 evaluate levels of expression and production of Env polypeptide. A modified synthetic coding sequence is presented as synthetic Env expression cassettes (Example 1, e.g., Tables 1A and 1B). The synthetic DNA fragments for Env were cloned into eucaryotic expression vectors  
15 described in Example 1 and in Section 2.1.3 above, including, a transient expression vector and CMV-promoter-based mammalian vectors. Corresponding wild-type sequences were cloned into the same vectors.

These vectors were then transfected into a several  
20 different cell types, including a variety of mammalian cell lines, (293, RD, COS-7, and CHO, cell lines available, for example, from the A.T.C.C.). The cell lines were cultured under appropriate conditions and the levels of gp120, gp140 and gp160 Env expression in  
25 supernatants were evaluated (Example 2). Env polypeptides include, but are not limited to, for example, native gp160, oligomeric gp140, monomeric gp120 as well as modified sequences of these polypeptides. The results of these assays demonstrated that expression of  
30 synthetic Env encoding sequences were significantly higher than corresponding wild-type sequences (Example 2; Tables 3 and 4).

Further, Western Blot analysis showed that cells containing the synthetic Env expression cassette produced



the expected protein (gp120, gp140 or gp160) at higher per-cell concentrations than cells containing the native expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production  
5 were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassettes of the present invention as compared to wild type.

Fractionation of the supernatants from mammalian cells transfected with the synthetic Env expression  
10 cassettes showed that it provides superior production of Env proteins, relative to the wild-type Env sequences (Examples 2 and 3).

Efficient expression of these Env-containing polypeptides in mammalian cell lines provides the  
15 following benefits: the Env polypeptides are free of baculovirus or other viral contaminants; production by established methods approved by the FDA; increased purity; greater yields (relative to native coding sequences); and a novel method of producing the Env-  
20 containing polypeptides in CHO cells which is less feasible in the absence of the increased expression obtained using the constructs of the present invention.

Exemplary cell lines (e.g., mammalian, yeast, insect, etc.) include those described above in Section  
25 2.1.3 for Gag-containing constructs. Further, appropriate vectors and control elements (e.g., promoters, enhancers, polyadenylation sequences, etc.) for any given cell type can be selected, as described above in Section 2.1.3, by one having ordinary skill in the art in view of the  
30 teachings of the present specification and information known in the art about expression vectors. In addition, the recombinantly expressed polypeptides from synthetic Env-encoding expression cassettes are typically isolated and purified from lysed cells or culture media, as

described above for Gag-encoding expression cassettes. An exemplary purification is described in Example 4 and shown in Figure 60.

5                    2.1.7            MODIFICATION OF HIV-1 TAT NUCLEIC ACID CODING  
SEQUENCES

Another aspect of the present invention is the generation of HIV-1 tat protein coding sequences, and related sequences, having improved expression relative to  
10 the corresponding wild-type sequence. Exemplary embodiments of the present invention are illustrated herein modifying the tat wild-type nucleotide sequence (SEQ ID NO:85, Figure 76) obtained from SF162 as described above. Exemplary synthetic tat constructs are  
15 shown in SEQ ID NO:87, which depicts a tat construct encoding a full-length tat polypeptide from strain SF162; SEQ ID NO:88, which depicts a tat construct encoding a tat polypeptide having the cystein residue at position 22 changed; and SEQ ID NO:89, which depicts a tat construct  
20 encoding the amino terminal portion of a tat polypeptide from strain SF162. The amino portion of the tat protein appears to contain many of the epitopes that induce an immune response. In addition, further modifications include replacement or deletion of the cystein residue at  
25 position 22, for example with a valine residue, an alanine residue or a glycine residue (SEQ ID Nos: 88 and 89, Figures 79 and 81), see, e.g., Caputo et al. (1996) *Gene Ther.* 3:235. In Figure 81, which depicts a tat construct encoding the amino terminal portion of a tat  
30 polypeptide, the nucleotides (nucleotides 64-66) encoding the cystein residues are underlined. The design and construction of suitable construct can be readily done using

the teachings of the present specification. As with Gag, pol, prot and Env, tat polypeptide coding sequences can be obtained from a variety of isolates (families, subtypes, etc.).

5        Modification of the tat polypeptide coding sequences result in improved expression relative to the wild-type coding sequences in a number of cell lines (e.g., mammalian, yeast, bacterial and insect cells). Tat polypeptide encoding sequences derived from these  
10        variants can be optimized and tested for improved expression in mammals by following the teachings of the present specification (see the Examples, in particular Example 2).

15        Various forms of the different embodiments of the invention, described herein, may be combined. For example, polynucleotides may be derived from the polynucleotide sequences of the present invention, including, but not limited to, coding sequences for Gag polypeptides, Env polypeptides, polymerase polypeptides,  
20        protease polypeptides, tat polypeptides, and reverse transcriptase polypeptides. Further, the polynucleotide coding sequences of the present invention may be combined into multi-cistronic expression cassettes where typically each coding sequence for each polypeptide is preceded by  
25        IRES sequences.

## 2.2 PRODUCTION OF VIRUS-LIKE PARTICLES AND USE OF THE CONSTRUCTS OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

30        The group-specific antigens (Gag) of human immunodeficiency virus type-1 (HIV-1) self-assemble into noninfectious virus-like particles (VLP) that are released from various eucaryotic cells by budding (reviewed by Freed, E.O., Virology 251:1-15, 1998). The

synthetic expression cassettes of the present invention provide efficient means for the production of HIV-Gag virus-like particles (VLPs) using a variety of different cell types, including, but not limited to, mammalian  
5 cells.

Viral particles can be used as a matrix for the proper presentation of an antigen entrapped or associated therewith to the immune system of the host. For example, U.S. Patent No. 4,722,840 describes hybrid particles  
10 comprised of a particle-forming fragment of a structural protein from a virus, such as a particle-forming fragment of hepatitis B virus (HBV) surface antigen (HBsAg), fused to a heterologous polypeptide. Tindle et al., *Virology* (1994) 200:547-557, describes the production and use of  
15 chimeric HBV core antigen particles containing epitopes of human papillomavirus (HPV) type 16 E7 transforming protein.

Adams et al., *Nature* (1987) 329:68-70, describes the recombinant production of hybrid HIVgpl20:Ty VLPs in  
20 yeast and Brown et al., *Virology* (1994) 198:477-488, the production of chimeric proteins consisting of the VP2 protein of human parvovirus B19 and epitopes from human herpes simplex virus type 1, as well as mouse hepatitis virus A59. Wagner et al., (*Virology* (1994) 200:162-175,  
25 Brand et al., *J. Virol. Meth.* (1995) 51:153-168; *Virology* (1996) 220:128-140) and Wolf, et al., (EP 0 449 116 A1, published 2 October 1991; WO 96/30523, published 3  
October 1996) describe the assembly of chimeric HIV-1 p55Gag particles. U.S. Patent No. 5,503,833 describes  
30 the use of rotavirus VP6 spheres for encapsulating and delivering therapeutic agents.

2.2.1 VLP PRODUCTION USING THE SYNTHETIC EXPRESSION  
CASSETTES OF THE PRESENT INVENTION

Experiments performed in support of the present invention have demonstrated that the synthetic expression  
5 cassettes of the present invention provide superior production of both protein and VLPs, relative to native coding sequences (Examples 7 and 15). Further, electron microscopic evaluation of VLP production (Examples 6 and 15, Figures 3A-B and 65A-F) showed that free and budding  
10 immature virus particles of the expected size were produced by cells containing the synthetic expression cassettes.

Using the synthetic expression cassettes of the present invention, rather than native coding sequences,  
15 for the production of virus-like particles provide several advantages. First, VLPs can be produced in enhanced quantity making isolation and purification of the VLPs easier. Second, VLPs can be produced in a variety of cell types using the synthetic expression  
20 cassettes, in particular, mammalian cell-lines can be used for VLP production, for example, CHO cells. Production using CHO cells provides (i) VLP formation; (ii) correct myristylation and budding; (iii) absence of non-mammalian cell contaminants (e.g., insect viruses  
25 and/or cells); and (iv) ease of purification. The synthetic expression cassettes of the present invention are also useful for enhanced expression in cell-types other than mammalian cell lines. For example, infection of insect cells with baculovirus vectors encoding the  
30 synthetic expression cassettes resulted in higher levels of total protein yield and higher levels of VLP production (relative to wild-type coding sequences). Further, the final product from insect cells infected with the baculovirus-Gag synthetic expression cassettes

consistently contained lower amounts of contaminating insect proteins than the final product when wild-type coding sequences were used (Examples).

VLPs can spontaneously form when the particle-forming polypeptide of interest is recombinantly expressed in an appropriate host cell. Thus, the VLPs produced using the synthetic expression cassettes of the present invention are conveniently prepared using recombinant techniques. As discussed below, the Gag polypeptide encoding synthetic expression cassettes of the present invention can include other polypeptide coding sequences of interest (for example, Env, tat, rev, HIV protease, HIV polymerase, HCV core; see, Example 1). Expression of such synthetic expression cassettes yields VLPs comprising the product of the synthetic expression cassette, as well as, the polypeptide of interest.

Once coding sequences for the desired particle-forming polypeptides have been isolated or synthesized, they can be cloned into any suitable vector or replicon for expression. Numerous cloning vectors are known to those of skill in the art, and the selection of an appropriate cloning vector is a matter of choice. See, generally, Ausubel et al, *supra* or Sambrook et al, *supra*. The vector is then used to transform an appropriate host cell. Suitable recombinant expression systems include, but are not limited to, bacterial, mammalian, baculovirus/insect, vaccinia, Semliki Forest virus (SFV), Alphaviruses (such as, Sindbis, Venezuelan Equine Encephalitis (VEE)), mammalian, yeast and Xenopus expression systems, well known in the art. Particularly preferred expression systems are mammalian cell lines, vaccinia, Sindbis, insect and yeast systems.

For example, a number of mammalian cell lines are known in the art and include immortalized cell lines

available from the American Type Culture Collection (A.T.C.C.), such as, but not limited to, Chinese hamster ovary (CHO) cells, 293 cells, HeLa cells, baby hamster kidney (BHK) cells, mouse myeloma (SB20), monkey kidney cells (COS), as well as others. Similarly, bacterial hosts such as *E. coli*, *Bacillus subtilis*, and *Streptococcus spp.*, will find use with the present expression constructs. Yeast hosts useful in the present invention include *inter alia*, *Saccharomyces cerevisiae*, *Candida albicans*, *Candida maltosa*, *Hansenula polymorpha*, *Kluyveromyces fragilis*, *Kluyveromyces lactis*, *Pichia guillerimondii*, *Pichia pastoris*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Insect cells for use with baculovirus expression vectors include, *inter alia*, *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. See, e.g., Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987). Fungal hosts include, for example, *Aspergillus*.

Viral vectors can be used for the production of particles in eucaryotic cells, such as those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. Additionally, a vaccinia based infection/transfection system, as described in Tomei et al., *J. Virol.* (1993) 67:4017-4026 and Selby et al., *J. Gen. Virol.* (1993) 74:1103-1113, will also find use with the present invention. In this system, cells are first infected *in vitro* with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the DNA of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus

recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. Alternately, T7 can be added as a purified protein or enzyme as in the "Progenitor" system (Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130). The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation product(s).

Depending on the expression system and host selected, the VLPS are produced by growing host cells transformed by an expression vector under conditions whereby the particle-forming polypeptide is expressed and VLPS can be formed. The selection of the appropriate growth conditions is within the skill of the art. If the VLPS are formed intracellularly, the cells are then disrupted, using chemical, physical or mechanical means, which lyse the cells yet keep the VLPS substantially intact. Such methods are known to those of skill in the art and are described in, e.g., *Protein Purification Applications: A Practical Approach*, (E.L.V. Harris and S. Angal, Eds., 1990).

The particles are then isolated (or substantially purified) using methods that preserve the integrity thereof, such as, by density gradient centrifugation, e.g., sucrose gradients, PEG-precipitation, pelleting, and the like (see, e.g., Kirnbauer et al. *J. Virol.* (1993) 67:6929-6936), as well as standard purification techniques including, e.g., ion exchange and gel filtration chromatography.

VLPS produced by cells containing the synthetic expression cassettes of the present invention can be used to elicit an immune response when administered to a subject. One advantage of the present invention is that VLPS can be produced by mammalian cells carrying the



synthetic expression cassettes at levels previously not possible. As discussed above, the VLPs can comprise a variety of antigens in addition to the Gag polypeptides (e.g., Env, tat, Gag-protease, Gag-polymerase, Gag-HCV-core). Purified VLPs, produced using the synthetic expression cassettes of the present invention, can be administered to a vertebrate subject, usually in the form of vaccine compositions. Combination vaccines may also be used, where such vaccines contain, for example, other subunit proteins derived from HIV or other organisms (e.g., env) or gene delivery vaccines encoding such antigens. Administration can take place using the VLPs formulated alone or formulated with other antigens. Further, the VLPs can be administered prior to, concurrent with, or subsequent to, delivery of the synthetic expression cassettes for DNA immunization (see below) and/or delivery of other vaccines. Also, the site of VLP administration may be the same or different as other vaccine compositions that are being administered. Gene delivery can be accomplished by a number of methods including, but are not limited to, immunization with DNA, alphavirus vectors, pox virus vectors, and vaccinia virus vectors.

VLP immune-stimulating (or vaccine) compositions can include various excipients, adjuvants, carriers, auxiliary substances, modulating agents, and the like. The immune stimulating compositions will include an amount of the VLP/antigen sufficient to mount an immunological response. An appropriate effective amount can be determined by one of skill in the art. Such an amount will fall in a relatively broad range that can be determined through routine trials and will generally be an amount on the order of about 0.1  $\mu$ g to about 1000  $\mu$ g,

more preferably about 1  $\mu$ g to about 300  $\mu$ g, of VLP/antigen.

A carrier is optionally present which is a molecule that does not itself induce the production of antibodies harmful to the individual receiving the composition.

Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus

particles. Examples of particulate carriers include those derived from polymethyl methacrylate polymers, as well as microparticles derived from poly(lactides) and poly(lactide-co-glycolides), known as PLG. See, e.g., Jeffery et al., *Pharm. Res.* (1993) 10:362-368; McGee JP, et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993. Such carriers are well known to those of ordinary skill in the art.

Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen may be conjugated to a bacterial toxoid, such as toxoid from diphtheria, tetanus, cholera, etc., as well as toxins derived from *E. coli*.

Such adjuvants include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc.; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59 (International Publication No. WO 90/14837), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated

into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA),  
(b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below)  
5 either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and  
(c) Rib<sup>i</sup>™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group  
10 consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particle generated therefrom such as  
15 ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (IL-1, IL-2, etc.), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), beta chemokines (MIP, 1-  
20 alpha, 1-beta Rantes, etc.); (6) detoxified mutants of a bacterial ADP-ribosylating toxin such as a cholera toxin (CT), a pertussis toxin (PT), or an *E. coli* heat-labile toxin (LT), particularly LT-K63 (where lysine is substituted for the wild-type amino acid at position 63)  
25 LT-R72 (where arginine is substituted for the wild-type amino acid at position 72), CT-S109 (where serine is substituted for the wild-type amino acid at position 109), and PT-K9/G129 (where lysine is substituted for the wild-type amino acid at position 9 and glycine  
30 substituted at position 129) (see, e.g., International Publication Nos. W093/13202 and W092/19265); and (7)

other substances that act as immunostimulating agents to enhance the effectiveness of the composition.

Muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acteyl-normuramyl-L-alanyl-D-isogluatme (nor-MDP), N-acetylmuramyl-L-alanyl-D-isogluatminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

Dosage treatment with the VLP composition may be a single dose schedule or a multiple dose schedule. A multiple dose schedule is one in which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals, chosen to maintain and/or reinforce the immune response, for example at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the potency of the modality, the vaccine delivery employed, the need of the subject and be dependent on the judgment of the practitioner.

If prevention of disease is desired (e.g., reduction of symptoms, recurrences or of disease progression), the antigen carrying VLPs are generally administered prior to primary infection with the pathogen of interest. If treatment is desired, e.g., the reduction of symptoms or recurrences, the VLP compositions are generally administered subsequent to primary infection.

#### 2.2.2 USING THE SYNTHETIC EXPRESSION CASSETTES OF THE PRESENT INVENTION TO CREATE PACKAGING CELL LINES

A number of viral based systems have been developed for use as gene transfer vectors for mammalian host cells. For example, retroviruses (in particular,

lentiviral vectors) provide a convenient platform for gene delivery systems. A coding sequence of interest (for example, a sequence useful for gene therapy applications) can be inserted into a gene delivery vector and packaged in retroviral particles using techniques known in the art. Recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described, including, for example, the following: (U.S. Patent No. 5,219,740; Miller et al. (1989) *Biotechniques* 7:980; Miller, A.D. (1990) *Human Gene Therapy* 1:5; Scarpa et al. (1991) *Virology* 180:849; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033; Boris-Lawrie et al. (1993) *Cur. Opin. Genet. Develop.* 3:102; GB 2200651; EP 0415731; EP 0345242; WO 89/02468; WO 89/05349; WO 89/09271; WO 90/02806; WO 90/07936; WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; WO 93/11230; WO 93/10218; WO 91/02805; in U.S. 5,219,740; U.S. 4,405,712; U.S. 4,861,719; U.S. 4,980,289 and U.S. 4,777,127; in U.S. Serial No. 07/800,921; and in Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53:83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci USA* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Sequences useful for gene therapy applications include, but are not limited to, the following. Factor VIII cDNA, including derivatives and deletions thereof (International Publication Nos. WO 96/21035, WO 97/03193, WO 97/03194, WO 97/03195, and WO 97/03191). Factor IX cDNA (Kurachi et al. (1982) *Proc. Natl. Acad. Sci. USA* 79:6461-6464). Factor V cDNA can be obtained from pMT2-V (Jenny (1987) *Proc. Natl. Acad. Sci. USA* 84:4846, A.T.C.C. Deposit No. 40515). A full-length factor V

cDNA, or a B domain deletion or B domain substitution thereof, can be used. B domain deletions of factor V, include those reported by Marquette (1995) *Blood* 86:3026 and Kane (1990) *Biochemistry* 29:6762. Antithrombin III cDNA (Prochownik (1983) *J. Biol. Chem.* 258:8389, A.T.C.C. Deposit No. 57224/57225). Protein C encoding cDNA (Foster (1984) *Proc. Natl. Acad. Sci. USA* 81:4766; Beckmann (1985) *Nucleic Acids Res.* 13:5233). Prothrombin cDNA can be obtained by restriction enzyme digestion of a published vector (Degen (1983) *Biochemistry* 22:2087). The endothelial cell surface protein, thrombomodulin, is a necessary cofactor for the normal activation of protein C by thrombin. A soluble recombinant form has been described (Parkinson (1990) *J. Biol. Chem.* 265:12602; Jackman (1987) *Proc. Natl. Acad. Sci. USA* 84:6425; Shirai (1988) *J. Biochem.* 103:281; Wen (1987) *Biochemistry* 26:4350; Suzuki (1987) *EMBO J.* 6:1891, A.T.C.C. Deposit No. 61348, 61349).

Many genetic diseases caused by inheritance of defective genes result in the failure to produce normal gene products, for example, thalassemia, phenylketonuria, Lesch-Nyhan syndrome, severe combined immunodeficiency (SCID), hemophilia A and B, cystic fibrosis, Duchenne's Muscular Dystrophy, inherited emphysema and familial hypercholesterolemia (Mulligan et al. (1993) *Science* 260:926; Anderson et al. (1992) *Science* 256:808; Friedman et al. (1989) *Science* 244:1275). Although genetic diseases may result in the absence of a gene product, endocrine disorders, such as diabetes and hypopituitarism, are caused by the inability of the gene to produce adequate levels of the appropriate hormone insulin and human growth hormone respectively.

In one aspect, gene therapy employing the constructs and methods of the present invention involves the

introduction of normal recombinant genes into T cells so that new or missing proteins are produced by the T cells after introduction or reintroduction thereof into a patient. A number of genetic diseases have been selected for treatment with gene therapy, including adenine deaminase deficiency, cystic fibrosis,  $\alpha_1$ -antitrypsin deficiency, Gaucher's syndrome, as well as non-genetic diseases.

In particular, Gaucher's syndrome is a genetic disorder characterized by a deficiency of the enzyme glucocerebrosidase. This enzyme deficiency leads to the accumulation of glucocerebroside in the lysosomes of all cells in the body. For a review see *Science* 256:794 (1992) and Scriver et al., *The Metabolic Basis of Inherited Disease*, 6th ed., vol. 2, page 1677). Thus, gene transfer vectors that express glucocerebrosidase can be constructed for use in the treatment of this disorder. Likewise, gene transfer vectors encoding lactase can be used in the treatment of hereditary lactose intolerance, those expressing AD can be used for treatment of ADA deficiency, and gene transfer vectors encoding  $\alpha_1$ -antitrypsin can be used to treat  $\alpha_1$ -antitrypsin deficiency. See Ledley, F.D. (1987) *J. Pediatrics* 110:157-174, Verma, I. (Nov. 1987) *Scientific American* pp. 68-84, and International Publication No. WO 95/27512 entitled "Gene Therapy Treatment for a Variety of Diseases and Disorders," for a description of gene therapy treatment of genetic diseases.

In still further embodiments of the invention, nucleotide sequences which can be incorporated into a gene transfer vector include, but are not limited to, proteins associated with enzyme-deficiency disorders, such as the cystic fibrosis transmembrane regulator (see, for example, U.S. Patent No. 5,240,846 and Larrick et al.

(1991) *Gene Therapy Applications of Molecular Biology*, Elsevier, New York and adenosine deaminase (ADA) (see U.S. Patent No. 5,399,346); growth factors, or an agonist or antagonist of a growth factor (Bandara et al. (1992) *DNA and Cell Biology*, 11:227); one or more tumor suppressor genes such as p53, Rb, or C-CAMI (Kleiner et al. (1995) *Cancer Research* 55:2831); a molecule that modulates the immune system of an organism, such as a HLA molecule (Nabel et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:11307); a ribozyme (Larsson et al. (1996) *Virology* 219:161); a peptide nucleic acid (Hirshman et al. (1996) *J. Invest. Med.* 44:347); an antisense molecule (Bordier et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:9383) which can be used to down-regulate the expression or synthesis of aberrant or foreign proteins, such as HIV proteins or a wide variety of oncogenes such as p53 (Hesketh, *The Oncogene Facts Book*, Academic Press, New York, (1995); a biopharmaceutical agent or antisense molecule used to treat HIV-infection, such as an inhibitor of p24 (Nakashima et al. (1994) *Nucleic Acids Res.* 22:5004); or reverse-transcriptase (see, Bordier, *supra*).

Other proteins of therapeutic interest can be expressed *in vivo* by gene transfer vectors using the methods of the invention. For instance sustained *in vivo* expression of tissue factor inhibitory protein (TFPI) is useful for treatment of conditions including sepsis and DIC and in preventing reperfusion injury. (See International Publications Nos. WO 93/24143, WO 93/25230 and WO 96/06637). Nucleic acid sequences encoding various forms of TFPI can be obtained, for example, as described in US Patent Nos. 4,966,852; 5,106,833; and 5,466,783, and incorporated into the gene transfer vectors described herein.



Erythropoietin (EPO) and leptin can also be expressed in vivo from genetically modified T cells according to the methods of the invention. For instance EPO is useful in gene therapy treatment of a variety of disorders including anemia (see International Publication No. WO 95/13376 entitled "Gene Therapy for Treatment of Anemia"). Sustained delivery of leptin by the methods of the invention is useful in treatment of obesity. See International Publication No. WO 96/05309 for a description of the leptin gene and the use thereof in the treatment of obesity.

A variety of other disorders can also be treated by the methods of the invention. For example, sustained in vivo systemic production of apolipoprotein E or apolipoprotein A from genetically modified T cells can be used for treatment of hyperlipidemia (see Breslow et al. (1994) *Biotechnology* 12:365). Sustained production of angiotensin receptor inhibitor (Goodfriend et al. (1996) *N. Engl. J. Med.* 334:1469) can be provided by the methods described herein. As yet an additional example, the long term in vivo systemic production of angiostatin is useful in the treatment of a variety of tumors. (See O'Reilly et al. (1996) *Nature Med.* 2:689).

In other embodiments, gene transfer vectors can be constructed to encode a cytokine or other immunomodulatory molecule. For example, nucleic acid sequences encoding native IL-2 and gamma-interferon can be obtained as described in US Patent Nos. 4,738,927 and 5,326,859, respectively, while useful muteins of these proteins can be obtained as described in U.S. Patent No. 4,853,332. Nucleic acid sequences encoding the short and long forms of mCSF can be obtained as described in US Patent Nos. 4,847,201 and 4,879,227, respectively. In particular aspects of the invention, retroviral vectors

expressing cytokine or immunomodulatory genes can be produced as described herein (for example, employing the packaging cell lines of the present invention) and in International Application No. PCT US 94/02951, entitled

5 "Compositions and Methods for Cancer Immunotherapy."

Examples of suitable immunomodulatory molecules for use herein include the following: IL-1 and IL-2 (Karupiah et al. (1990) *J. Immunology* 144:290-298, Weber et al. (1987) *J. Exp. Med.* 166:1716-1733, Gansbacher et al. (1990) *J. Exp. Med.* 172:1217-1224, and U.S. Patent No. 4,738,927); IL-3 and IL-4 (Tepper et al. (1989) *Cell* 57:503-512, Golumbek et al. (1991) *Science* 254:713-716, and U.S. Patent No. 5,017,691); IL-5 and IL-6 (Brakenhof et al. (1987) *J. Immunol.* 139:4116-4121, and

15 International Publication No. WO 90/06370); IL-7 (U.S. Patent No. 4,965,195); IL-8, IL-9, IL-10, IL-11, IL-12, and IL-13 (*Cytokine Bulletin*, Summer 1994); IL-14 and IL-15; alpha interferon (Finter et al. (1991) *Drugs* 42:749-765, U.S. Patent Nos. 4,892,743 and 4,966,843,

20 International Publication No. WO 85/02862, Nagata et al. (1980) *Nature* 284:316-320, Familletti et al. (1981) *Methods in Enz.* 78:387-394, Twu et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2046-2050, and Faktor et al. (1990) *Oncogene* 5:867-872); beta-interferon (Seif et al. (1991) *J. Virol.* 65:664-671); gamma-interferons (Radford et al. (1991) *The American Society of Hepatology* 20082015, Watanabe et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:9456-9460, Gansbacher et al. (1990) *Cancer Research* 50:7820-7825, Maio et al. (1989) *Can. Immunol.*

25 *Immunother.* 30:34-42, and U.S. Patent Nos. 4,762,791 and 4,727,138); G-CSF (U.S. Patent Nos. 4,999,291 and 4,810,643); GM-CSF (International Publication No. WO 85/04188); tumor necrosis factors (TNFs) (Jayaraman et al. (1990) *J. Immunology* 144:942-951); CD3 (Krissanen et

30

al. (1987) *Immunogenetics* 26:258-266); ICAM-1 (Altman et al. (1989) *Nature* 338:512-514, Simmons et al. (1988) *Nature* 331:624-627); ICAM-2, LFA-1, LFA-3 (Wallner et al. (1987) *J. Exp. Med.* 166:923-932); MHC class I molecules, 5 MHC class II molecules, B7.1-.3,  $\beta_2$ -microglobulin (Parnes et al. (1981) *Proc. Natl. Acad. Sci. USA* 78:2253-2257); chaperones such as calnexin; and MHC-linked transporter proteins or analogs thereof (Powis et al. (1991) *Nature* 354:528-531). Immunomodulatory factors may also be 10 agonists, antagonists, or ligands for these molecules. For example, soluble forms of receptors can often behave as antagonists for these types of factors, as can mutated forms of the factors themselves.

Nucleic acid molecules that encode the above- 15 described substances, as well as other nucleic acid molecules that are advantageous for use within the present invention, may be readily obtained from a variety of sources, including, for example, depositories such as the American Type Culture Collection, or from commercial 20 sources such as British Bio-Technology Limited (Cowley, Oxford England). Representative examples include BBG 12 (containing the GM-CSF gene coding for the mature protein of 127 amino acids), BBG 6 (which contains sequences encoding gamma interferon), A.T.C.C. Deposit No. 39656 25 (which contains sequences encoding TNF), A.T.C.C. Deposit No. 20663 (which contains sequences encoding alpha-interferon), A.T.C.C. Deposit Nos. 31902, 31902 and 39517 (which contain sequences encoding beta-interferon), A.T.C.C. Deposit No. 67024 (which contains a sequence 30 which encodes Interleukin-1b), A.T.C.C. Deposit Nos. 39405, 39452, 39516, 39626 and 39673 (which contain sequences encoding Interleukin-2), A.T.C.C. Deposit Nos. 59399, 59398, and 67326 (which contain sequences encoding Interleukin-3), A.T.C.C. Deposit No. 57592 (which

contains sequences encoding Interleukin-4), A.T.C.C. Deposit Nos. 59394 and 59395 (which contain sequences encoding Interleukin-5), and A.T.C.C. Deposit No. 67153 (which contains sequences encoding Interleukin-6).

5        Plasmids containing cytokine genes or immunomodulatory genes (International Publication Nos. WO 94/02951 and WO 96/21015) can be digested with appropriate restriction enzymes, and DNA fragments containing the particular gene of interest can be inserted into a gene  
10       transfer vector using standard molecular biology techniques. (See, e.g., Sambrook et al., *supra.*, or Ausubel et al. (eds) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience).

Exemplary hormones, growth factors and other  
15       proteins which are useful for long term expression are described, for example, in European Publication No. 0437478B1, entitled "Cyclodextrin-Peptide Complexes." Nucleic acid sequences encoding a variety of hormones can be used, including those encoding human growth hormone,  
20       insulin, calcitonin, prolactin, follicle stimulating hormone (FSH), luteinizing hormone (LH), human chorionic gonadotropin (HCG), and thyroid stimulating hormone (TSH). A variety of different forms of IGF-1 and IGF-2 growth factor polypeptides are also well known the art  
25       and can be incorporated into gene transfer vectors for long term expression *in vivo*. See, e.g., European Patent No. 0123228B1, published for grant September 19, 1993, entitled "Hybrid DNA Synthesis of Mature Insulin-like Growth Factors." As an additional example, the long term  
30       *in vivo* expression of different forms of fibroblast growth factor can also be effected employing the compositions and methods of invention. See, e.g., U.S. Patent Nos. 5,464,774, 5,155,214, and 4,994,559 for a description of different fibroblast growth factors.

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene  
5 from a vector known to include the same. For example, plasmids which contain sequences that encode altered cellular products may be obtained from a depository such as the A.T.C.C., or from commercial sources. Plasmids containing the nucleotide sequences of interest can be  
10 digested with appropriate restriction enzymes, and DNA fragments containing the nucleotide sequences can be inserted into a gene transfer vector using standard molecular biology techniques.

Alternatively, cDNA sequences for use with the  
15 present invention may be obtained from cells which express or contain the sequences, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain and isolate DNA.  
20 Briefly, mRNA from a cell which expresses the gene of interest can be reverse transcribed with reverse transcriptase using oligo-dT or random primers. The single stranded cDNA may then be amplified by PCR (see U.S. Patent Nos. 4,683,202, 4,683,195 and 4,800,159, see  
25 also *PCR Technology: Principles and Applications for DNA Amplification*, Erlich (ed.), Stockton Press, 1989)) using oligonucleotide primers complementary to sequences on either side of desired sequences.

The nucleotide sequence of interest can also be  
30 produced synthetically, rather than cloned, using a DNA synthesizer (e.g., an Applied Biosystems Model 392 DNA Synthesizer, available from ABI, Foster City, California). The nucleotide sequence can be designed with the appropriate codons for the expression product

desired. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge (1981) *Nature* 292:756; Nambair et al. (1984) *Science* 223:1299; Jay et al. (1984) *J. Biol. Chem.* 259:6311.

The synthetic expression cassettes of the present invention can be employed in the construction of packaging cell lines for use with retroviral vectors.

One type of retrovirus, the murine leukemia virus, or "MLV", has been widely utilized for gene therapy applications (see generally Mann et al. (*Cell* 33:153, 1993), Cane and Mulligan (*Proc. Nat'l. Acad. Sci. USA* 81:6349, 1984), and Miller et al., *Human Gene Therapy* 1:5-14, 1990).

Lentiviral vectors typically, comprise a 5' lentiviral LTR, a tRNA binding site, a packaging signal, a promoter operably linked to one or more genes of interest, an origin of second strand DNA synthesis and a 3' lentiviral LTR, wherein the lentiviral vector contains a nuclear transport element. The nuclear transport element may be located either upstream (5') or downstream (3') of a coding sequence of interest. Within certain embodiments, the nuclear transport element is not RRE. Within one embodiment the packaging signal is an extended packaging signal. Within other embodiments the promoter is a tissue specific promoter, or, alternatively, a promoter such as CMV. Within other embodiments, the lentiviral vector further comprises an internal ribosome entry site.

A wide variety of lentiviruses may be utilized within the context of the present invention, including for example, lentiviruses selected from the group consisting of HIV, HIV-1, HIV-2, FIV and SIV.

In one embodiment of the present invention synthetic Env and/or Gag-polymerase expression cassettes are provided comprising a promoter and a sequence encoding synthetic Gag-polymerase (SEQ ID NO:6) and at least one  
5 of vpr, vpu, nef or vif, wherein the promoter is operably linked to Gag-polymerase and vpr, vpu, nef or vif.

Within yet another aspect of the invention, host cells (e.g., packaging cell lines) are provided which contain any of the expression cassettes described herein.  
10 For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic Env and/or Gag-polymerase, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding Env and/or Gag-  
15 polymerase. Packaging cell lines may further comprise a promoter and a sequence encoding tat, rev, or an envelope, wherein the promoter is operably linked to the sequence encoding tat, rev, or, the envelope. The packaging cell line may further comprise a sequence  
20 encoding any one or more of nef, vif, vpu or vpr.

In one embodiment, the expression cassette (carrying, for example, the synthetic Env, synthetic tat and/or synthetic Gag-polymerase) is stably integrated. The packaging cell line, upon introduction of a  
25 lentiviral vector, typically produces viral particles. The promoter regulating expression of the synthetic expression cassette may be inducible. Typically, the packaging cell line, upon introduction of a lentiviral vector, produces viral particles that are essentially  
30 free of replication competent virus.

Packaging cell lines are provided comprising an expression cassette which directs the expression of a synthetic Env (or Gag-polymerase) gene, an expression cassette which directs the expression of a Gag (or Env)

gene optimized for expression (e.g., Andre, S., et al.,  
*Journal of Virology* 72(2):1497-1503, 1998; Haas, J., et  
al., *Current Biology* 6(3):315-324, 1996). A lentiviral  
vector is introduced into the packaging cell line to  
5 produce a vector particle producing cell line.

As noted above, lentiviral vectors can be designed  
to carry or express a selected gene(s) or sequences of  
interest. Lentiviral vectors may be readily constructed  
from a wide variety of lentiviruses (see RNA Tumor  
10 Viruses, Second Edition, Cold Spring Harbor Laboratory,  
1985). Representative examples of lentiviruses included  
HIV, HIV-1, HIV-2, FIV and SIV. Such lentiviruses may  
either be obtained from patient isolates, or, more  
preferably, from depositories or collections such as the  
15 American Type Culture Collection, or isolated from known  
sources using available techniques.

Portions of the lentiviral gene delivery vectors (or  
vehicles) may be derived from different viruses. For  
example, in a given recombinant lentiviral vector, LTRs  
20 may be derived from an HIV, a packaging signal from SIV,  
and an origin of second strand synthesis from HrV-2.  
Lentiviral vector constructs may comprise a 5' lentiviral  
LTR, a tRNA binding site, a packaging signal, one or more  
heterologous sequences, an origin of second strand DNA  
25 synthesis and a 3' LTR, wherein said lentiviral vector  
contains a nuclear transport element that is not RRE.

Briefly, Long Terminal Repeats ("LTRs") are  
subdivided into three elements, designated U5, R and U3.  
These elements contain a variety of signals which are  
30 responsible for the biological activity of a retrovirus,  
including for example, promoter and enhancer elements  
which are located within U3. LTRs may be readily  
identified in the provirus (integrated DNA form) due to  
their precise duplication at either end of the genome.



As utilized herein, a 5' LTR should be understood to include a 5' promoter element and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector. The 3' LTR should be understood to include a polyadenylation signal, and sufficient LTR sequence to allow reverse transcription and integration of the DNA form of the vector.

The tRNA binding site and origin of second strand DNA synthesis are also important for a retrovirus to be biologically active, and may be readily identified by one of skill in the art. For example, retroviral tRNA binds to a tRNA binding site by Watson-Crick base pairing, and is carried with the retrovirus genome into a viral particle. The tRNA is then utilized as a primer for DNA synthesis by reverse transcriptase. The tRNA binding site may be readily identified based upon its location just downstream from the 5'LTR. Similarly, the origin of second strand DNA synthesis is, as its name implies, important for the second strand DNA synthesis of a retrovirus. This region, which is also referred to as the poly-purine tract, is located just upstream of the 3'LTR.

In addition to a 5' and 3' LTR, tRNA binding site, and origin of second strand DNA synthesis, recombinant retroviral vector constructs may also comprise a packaging signal, as well as one or more genes or coding sequences of interest. In addition, the lentiviral vectors have a nuclear transport element which, in preferred embodiments is not RRE. Representative examples of suitable nuclear transport elements include the element in Rous sarcoma virus (Ogert, et al., *J ViroL* 70, 3834-3843, 1996), the element in Rous sarcoma virus (Liu & Mertz, *Genes & Dev.*, 9, 1766-1789, 1995) and the element in the genome of simian retrovirus type I

(Zolotukhin, et al., *J Virol.* 68, 7944-7952, 1994).

Other potential elements include the elements in the histone gene (Kedes, *Annu. Rev. Biochem.* 48, 837-870, 1970), the  $\alpha$ -interferon gene (Nagata et al., *Nature* 287, 401-408, 1980), the  $\beta$ -adrenergic receptor gene (Koilkka, et al., *Nature* 329, 75-79, 1987), and the c-Jun gene (Hattorie, et al., *Proc. Natl. Acad. Sci. USA* 85, 9148-9152, 1988).

Recombinant lentiviral vector constructs typically lack both *Gag*-polymerase and *env* coding sequences. Recombinant lentiviral vector typically contain less than 20, preferably 15, more preferably 10, and most preferably 8 consecutive nucleotides found in *Gag*-polymerase or *env* genes. One advantage of the present invention is that the synthetic *Gag*-polymerase expression cassettes, which can be used to construct packaging cell lines for the recombinant retroviral vector constructs, have little homology to wild-type *Gag*-polymerase sequences and thus considerably reduce or eliminate the possibility of homologous recombination between the synthetic and wild-type sequences.

Lentiviral vectors may also include tissue-specific promoters to drive expression of one or more genes or sequences of interest. For example, lentiviral vector particles of the invention can contain a liver specific promoter to maximize the potential for liver specific expression of the exogenous DNA sequence contained in the vectors. Preferred liver specific promoters include the hepatitis B X-gene promoter and the hepatitis B core protein promoter. These liver specific promoters are preferably employed with their respective enhancers. The enhancer element can be linked at either the 5' or the 3' end of the nucleic acid encoding the sequences of interest. The hepatitis B X gene promoter and its

enhancer can be obtained from the viral genome as a 332 base pair *EcoRV*-*NcoI* DNA fragment employing the methods described in Twu, et al., *J Virol.* 61:3448-3453, 1987.

The hepatitis B core protein promoter can be obtained

5 from the viral genome as a 584 base pair *BamHI*-*BglIII* DNA fragment employing the methods described in Gerlach, et al., *Virol* 189:59-66, 1992. It may be necessary to remove the negative regulatory sequence in the *BamHI*-*BglIII* fragment prior to inserting it. Other liver  
10 specific promoters include the AFP (alpha fetal protein) gene promoter and the albumin gene promoter, as disclosed in EP Patent Publication 0 415 731, the -1 antitrypsin gene promoter, as disclosed in Rettenger, et al., *Proc. Natl. Acad. Sci.* 91:1460-1464, 1994, the fibrinogen  
15 gene promoter, the APO-A1 (Apolipoprotein A1) gene promoter, and the promoter genes for liver transference enzymes such as, for example, SGOT, SGPT and glutamyle transferase. See also PCT Patent Publications WO 90/07936 and WO 91/02805 for a description of the use of  
20 liver specific promoters in lentiviral vector particles.

Lentiviral vector constructs may be generated such that more than one gene of interest is expressed. This may be accomplished through the use of di- or oligo-cistronic cassettes (e.g., where the coding regions are  
25 separated by 80 nucleotides or less, see generally Levin et al., *Gene* 108:167-174, 1991), or through the use of Internal Ribosome Entry Sites ("IRES").

Packaging cell lines suitable for use with the above described recombinant retroviral vector constructs may be  
30 readily prepared given the disclosure provided herein. Briefly, the parent cell line from which the packaging cell line is derived can be selected from a variety of

mammalian cell lines, including for example, 293, RD, COS-7, CHO, BHK, VERO, HT1080, and myeloma cells.

After selection of a suitable host cell for the generation of a packaging cell line, one or more  
5 expression cassettes are introduced into the cell line in order to complement or supply in trans components of the vector which have been deleted.

Representative examples of suitable expression cassettes have been described herein and include  
10 synthetic Env, tat, Gag, synthetic Gag-protease, synthetic Gag-reverse transcriptase and synthetic Gag-polymerase expression cassettes, which comprise a promoter and a sequence encoding, e.g., Env, tat, or Gag-polymerase and at least one of vpr, vpu, nef or vif,  
15 wherein the promoter is operably linked to Env, tat or Gag-polymerase and vpr, vpu, nef or vif. As described above, optimized Env, Gag and/or tat coding sequences may also be utilized in various combinations in the generation of packaging cell lines.

20 Utilizing the above-described expression cassettes, a wide variety of packaging cell lines can be generated. For example, within one aspect packaging cell line are provided comprising an expression cassette that comprises a sequence encoding synthetic HIV (e.g., Gag, Env, tat, Gag-polymerase, Gag-reverse transcriptase or Gag-  
25 protease) polypeptide, and a nuclear transport element, wherein the promoter is operably linked to the sequence encoding the HIV polypeptide. Within other aspects, packaging cell lines are provided comprising a promoter and a sequence encoding Gag, tat, rev, or an envelope  
30 (e.g., HIV env), wherein the promoter is operably linked to the sequence encoding Gag, tat, rev, or, the envelope. Within further embodiments, the packaging cell line may comprise a sequence encoding any one or more of nef, vif,

vpu or vpr. For example, the packaging cell line may contain only nef, vif, vpu, or vpr alone, nef and vif, nef and vpu, nef and vpr, vif and vpu, vif and vpr, vpu and vpr, nef vif and vpu, nef vif and vpr, nef vpu and vpr, vif vpu and vpr, or, all four of nef vif vpu and vpr.

In one embodiment, the expression cassette is stably integrated. Within another embodiment, the packaging cell line, upon introduction of a lentiviral vector, produces particles. Within further embodiments the promoter is inducible. Within certain preferred embodiments of the invention, the packaging cell line, upon introduction of a lentiviral vector, produces particles that are free of replication competent virus.

The synthetic cassettes containing optimized coding sequences are transfected into a selected cell line. Transfected cells are selected that (i) carry, typically, integrated, stable copies of the Gag, Pol, and Env coding sequences, and (ii) are expressing acceptable levels of these polypeptides (expression can be evaluated by methods known in the prior art, e.g., see Examples 1-4). The ability of the cell line to produce VLPs may also be verified (Examples 6, 7 and 15).

A sequence of interest is constructed into a suitable viral vector as discussed above. This defective virus is then transfected into the packaging cell line. The packaging cell line provides the viral functions necessary for producing virus-like particles into which the defective viral genome, containing the sequence of interest, are packaged. These VLPs are then isolated and can be used, for example, in gene delivery or gene therapy.

Further, such packaging cell lines can also be used to produce VLPs alone, which can, for example, be used as

adjuvants for administration with other antigens or in vaccine compositions. Also, co-expression of a selected sequence of interest encoding a polypeptide (for example, an antigen) in the packaging cell line can also result in the entrapment and/or association of the selected polypeptide in/with the VLPs.

### 2.3 DNA IMMUNIZATION AND GENE DELIVERY

A variety of polypeptide antigens can be used in the practice of the present invention. Polypeptide antigens can be included in DNA immunization constructs containing, for example, any of the synthetic expression cassettes described herein fused in-frame to a coding sequence for the polypeptide antigen, where expression of the construct results in VLPs presenting the antigen of interest. Antigens can be derived from a wide variety of viruses, bacteria, fungi, plants, protozoans and other parasites. For example, the present invention will find use for stimulating an immune response against a wide variety of proteins from the herpesvirus family, including proteins derived from herpes simplex virus (HSV) types 1 and 2, such as HSV-1 and HSV-2 gB, gD, gH, VP16 and VP22; antigens derived from varicella zoster virus (VZV), Epstein-Barr virus (EBV) and cytomegalovirus (CMV) including CMV gB and gH; and antigens derived from other human herpesviruses such as HHV6 and HHV7. (See, e.g. Chee et al., *Cytomegaloviruses* (J.K. McDougall, ed., Springer-Verlag 1990) pp. 125-169, for a review of the protein coding content of cytomegalovirus; McGeoch et al., *J. Gen. Virol.* (1988) 69:1531-1574, for a discussion of the various HSV-1 encoded proteins; U.S. Patent No. 5,171,568 for a discussion of HSV-1 and HSV-2 gB and gD proteins and the genes encoding therefore; Baer et al., *Nature* (1984) 310:207-211, for the identification of

protein coding sequences in an EBV genome; and Davison and Scott, *J. Gen. Virol.* (1986) 67:1759-1816, for a review of VZV.)

5 Additionally, immune responses to antigens from the hepatitis family of viruses, including hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the delta hepatitis virus (HDV), hepatitis E virus (HEV), and hepatitis G virus, can also be stimulated using the constructs of the present invention. By way of example,  
10 the HCV genome encodes several viral proteins, including E1 (also known as E) and E2 (also known as E2/NSI), which will find use with the present invention (see, Houghton et al. *Hepatology* (1991) 14:381-388, for a discussion of HCV proteins, including E1 and E2). The  $\delta$ -antigen from  
15 HDV can also be used (see, e.g., U.S. Patent No. 5,389,528, for a description of the  $\delta$ -antigen).

Similarly, influenza virus is another example of a virus for which the present invention will be particularly useful. Specifically, the envelope  
20 glycoproteins HA and NA of influenza A are of particular interest for generating an immune response. Numerous HA subtypes of influenza A have been identified (Kawaoka et al., *Virology* (1990) 179:759-767; Webster et al. "Antigenic variation among type A influenza viruses," p.  
25 127-168. In: P. Palese and D.W. Kingsbury (ed.), *Genetics of influenza viruses*. Springer-Verlag, New York).

Other antigens of particular interest to be used in the practice of the present invention include antigens and polypeptides derived therefrom from human  
30 papillomavirus (HPV), such as one or more of the various early proteins including E6 and E7; tick-borne encephalitis viruses; and HIV-1 (also known as HTLV-III, LAV, ARV, etc.), including, but not limited to, antigens such as gp120, gp41, gp160, Gag and pol from a variety of

isolates including, but not limited to, HIV<sub>IIIB</sub>, HIV<sub>SF2</sub>, HIV-1<sub>SF162</sub>, HIV-1<sub>SF170</sub>, HIV<sub>LAV</sub>, HIV<sub>LAI</sub>, HIV<sub>MN</sub>, HIV-1<sub>CM235</sub>, HIV-1<sub>US4</sub>, other HIV-1 strains from diverse subtypes (e.g., subtypes, A through G, and O), HIV-2 strains and diverse subtypes (e.g., HIV-2<sub>UC1</sub> and HIV-2<sub>UC2</sub>). See, e.g., Myers, et al., Los Alamos Database, Los Alamos National Laboratory, Los Alamos, New Mexico; Myers, et al., *Human Retroviruses and Aids*, 1990, Los Alamos, New Mexico: Los Alamos National Laboratory.

Proteins derived from other viruses will also find use in the claimed methods, such as without limitation, proteins from members of the families Picornaviridae (e.g., polioviruses, etc.); Caliciviridae; Togaviridae (e.g., rubella virus, dengue virus, etc.); Flaviviridae; Coronaviridae; Reoviridae; Birnaviridae; Rhabdoviridae (e.g., rabies virus, etc.); Filoviridae; Paramyxoviridae (e.g., mumps virus, measles virus, respiratory syncytial virus, etc.); Orthomyxoviridae (e.g., influenza virus types A, B and C, etc.); Bunyaviridae; Arenaviridae; Retroviridae, e.g., HTLV-I; HTLV-II; HIV-1; HIV-2; simian immunodeficiency virus (SIV) among others. See, e.g. *Virology*, 3rd Edition (W.K. Joklik ed. 1988); *Fundamental Virology*, 2nd Edition (B.N. Fields and D.M. Knipe, eds. 1991; *Virology*, 3rd Edition (Fields, BN, DM Knipe, PM Howley, Editors, 1996, Lippincott-Raven, Philadelphia, PA) for a description of these and other viruses.

Particularly preferred bacterial antigens are derived from organisms that cause diphtheria, tetanus, pertussis, meningitis, and other pathogenic states, including, without limitation, antigens derived from *Corynebacterium diphtheriae*, *Clostridium tetani*, *Bordetella pertussis*, *Neisseria meningitidis*, including serotypes *Meningococcus* A, B, C, Y and WI35 (MenA, B, C, Y and WI35), *Haemophilus influenza* type B (Hib), and



*Helicobacter pylori*. Examples of parasitic antigens include those derived from organisms causing malaria, tuberculosis, and Lyme disease.

Furthermore, the methods described herein provide  
5 means for treating a variety of malignant cancers. For example, the system of the present invention can be used to enhance both humoral and cell-mediated immune responses to particular proteins specific to a cancer in question, such as an activated oncogene, a fetal antigen,  
10 or an activation marker. Such tumor antigens include any of the various MAGEs (melanoma associated antigen E), including MAGE 1, 2, 3, 4, etc. (Boon, T. *Scientific American* (March 1993):82-89); any of the various tyrosinases; MART 1 (melanoma antigen recognized by T  
15 cells), mutant ras; mutant p53; p97 melanoma antigen; CEA (carcinoembryonic antigen), among others.

DNA immunization using synthetic expression cassettes of the present invention has been demonstrated to be efficacious (Examples 8 and 10-12). Animals were  
20 immunized with both the synthetic expression cassette and the wild type expression cassette. The results of the immunizations with plasmid-DNAs showed that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression  
25 cassettes. Also, the second boost immunization induced a secondary immune response, for example after two to eight weeks. Further, the results of CTL assays showed increased potency of synthetic expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by  
30 DNA immunization.

It is readily apparent that the subject invention can be used to mount an immune response to a wide variety of antigens and hence to treat or prevent a large number of diseases.

2.3.1 DELIVERY OF THE SYNTHETIC EXPRESSION CASSETTES OF THE  
PRESENT INVENTION

Polynucleotide sequences coding for the above-described molecules can be obtained using recombinant  
5 methods, such as by screening cDNA and genomic libraries from cells expressing the gene, or by deriving the gene from a vector known to include the same. The sequences can be analyzed by conventional sequencing techniques. Furthermore, the desired gene can be isolated directly  
10 from cells and tissues containing the same, using standard techniques, such as phenol extraction and PCR of cDNA or genomic DNA. See, e.g., Sambrook et al., *supra*, for a description of techniques used to obtain, isolate and sequence DNA. Once the sequence is known, the gene  
15 of interest can also be produced synthetically, rather than cloned. The nucleotide sequence can be designed with the appropriate codons for the particular amino acid sequence desired. In general, one will select preferred codons for the intended host in which the sequence will  
20 be expressed. The complete sequence is assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete coding sequence. See, e.g., Edge, *Nature* (1981) 292:756; Nambair et al., *Science* (1984) 223:1299; Jay et al., *J. Biol. Chem.* (1984) 259:6311; Stemmer, W.P.C., (1995) *Gene*  
25 164:49-53.

Next, the gene sequence encoding the desired antigen can be inserted into a vector containing a synthetic expression cassette of the present invention (e.g., see  
30 Example 1 for construction of various exemplary synthetic expression cassette). The antigen is inserted into the synthetic coding sequence such that when the combined sequence is expressed it results in the production of VLPs comprising the polypeptide and/or the antigen of

interest. Insertions can be made within the Gag coding sequence or at either end of the coding sequence (5', amino terminus of the expressed polypeptide; or 3', carboxy terminus of the expressed polypeptide -- e.g., see Example 1) (Wagner, R., et al., *Arch Virol.* 127:117-137, 1992; Wagner, R., et al., *Virology* 200:162-175, 1994; Wu, X., et al., *J. Virol.* 69(6):3389-3398, 1995; Wang, C-T., et al., *Virology* 200:524-534, 1994; Chazal, N., et al., *Virology* 68(1):111-122, 1994; Griffiths, J.C., et al., *J. Virol.* 67(6):3191-3198, 1993; Reicin, A.S., et al., *J. Virol.* 69(2):642-650, 1995).

Up to 50% of the coding sequences of p55Gag can be deleted without affecting the assembly to virus-like particles and expression efficiency (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Gamier, L., et al., *J Virol* 72(6):4667-4677, 1998; Zhang, Y., et al., *J Virol* 72(3):1782-1789, 1998; Wang, C., et al., *J Virol* 72(10):7950-7959, 1998). In one embodiment of the present invention, immunogenicity of the high level expressing synthetic p55GagMod and p55GagProtMod expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted, mutated or truncated regions of p55GagMod sequence. In another embodiment of the present invention, immunogenicity of the high level expressing synthetic Env expression cassettes can be increased by the insertion of different structural or non-structural HIV antigens, multiepitope cassettes, or cytokine sequences into deleted regions of gp120Mod, gp140Mod or gp160Mod sequences. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length modified Env

sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or higher immunogenicity of the expression product. Such deletions may be generated following the teachings of the present invention and information available to one of ordinary skill in the art. One possible advantage of this approach, relative to using full-length Env, Gag or Tat sequences fused to heterologous polypeptides, can be higher expression/secretion efficiency and/or immunogenicity of the expression product.

When sequences are added to the amino terminal end of Gag (for example, when using the synthetic p55GagMod expression cassette of the present invention), the polynucleotide can contain coding sequences at the 5' end that encode a signal for addition of a myristic moiety to the Gag-containing polypeptide (e.g., sequences that encode Met-Gly).

The ability of Gag-containing polypeptide constructs to form VLPs can be empirically determined following the teachings of the present specification.

HIV polypeptide/antigen synthetic expression cassettes include control elements operably linked to the coding sequence, which allow for the expression of the gene *in vivo* in the subject species. For example, typical promoters for mammalian cell expression include the SV40 early promoter, a CMV promoter such as the CMV immediate early promoter, the mouse mammary tumor virus LTR promoter, the adenovirus major late promoter (Ad MLP), and the herpes simplex virus promoter, among others. Other nonviral promoters, such as a promoter derived from the murine metallothionein gene, will also find use for mammalian expression. Typically, transcription termination and polyadenylation sequences will also be present, located 3' to the translation stop

codon. Preferably, a sequence for optimization of initiation of translation, located 5' to the coding sequence, is also present. Examples of transcription terminator/polyadenylation signals include those derived from SV40, as described in Sambrook et al., *supra*, as well as a bovine growth hormone terminator sequence.

Enhancer elements may also be used herein to increase expression levels of the mammalian constructs. Examples include the SV40 early gene enhancer, as described in Dijkema et al., *EMBO J.* (1985) 4:761, the enhancer/promoter derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus, as described in Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777 and elements derived from human CMV, as described in Boshart et al., *Cell* (1985) 41:521, such as elements included in the CMV intron A sequence.

Furthermore, plasmids can be constructed which include a chimeric antigen-coding gene sequences, encoding, e.g., multiple antigens/epitopes of interest, for example derived from a single or from more than one viral isolate.

Typically the antigen coding sequences precede or follow the synthetic coding sequences and the chimeric transcription unit will have a single open reading frame encoding both the antigen of interest and the synthetic Gag coding sequences. Alternatively, multi-cistronic cassettes (e.g., bi-cistronic cassettes) can be constructed allowing expression of multiple antigens from a single mRNA using the EMCV IRES, or the like. Lastly, antigens can be encoded on separate transcripts from independent promoters on a single plasmid or other vector.

Once complete, the constructs are used for nucleic acid immunization or the like using standard gene

delivery protocols. Methods for gene delivery are known in the art. See, e.g., U.S. Patent Nos. 5,399,346, 5,580,859, 5,589,466. Genes can be delivered either directly to the vertebrate subject or, alternatively, delivered *ex vivo*, to cells derived from the subject and the cells reimplanted in the subject.

A number of viral based systems have been developed for gene transfer into mammalian cells. For example, retroviruses provide a convenient platform for gene delivery systems. Selected sequences can be inserted into a vector and packaged in retroviral particles using techniques known in the art. The recombinant virus can then be isolated and delivered to cells of the subject either *in vivo* or *ex vivo*. A number of retroviral systems have been described (U.S. Patent No. 5,219,740; Miller and Rosman, *BioTechniques* (1989) 7:980-990; Miller, A.D., *Human Gene Therapy* (1990) 1:5-14; Scarpa et al., *Virology* (1991) 180:849-852; Burns et al., *Proc. Natl. Acad. Sci. USA* (1993) 90:8033-8037; and Boris-Lawrie and Temin, *Cur. Opin. Genet. Develop.* (1993) 3:102-109.

A number of adenovirus vectors have also been described. Unlike retroviruses which integrate into the host genome, adenoviruses persist extrachromosomally thus minimizing the risks associated with insertional mutagenesis (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K.L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476).

Additionally, various adeno-associated virus (AAV) vector systems have been developed for gene delivery.

AAV vectors can be readily constructed using techniques well known in the art. See, e.g., U.S. Patent Nos.

5,173,414 and 5,139,941; International Publication Nos. WO 92/01070 (published 23 January 1992) and WO 93/03769

- 5 (published 4 March 1993); Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines* 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B.J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* 10 (1992) 158:97-129; Kotin, R.M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875.

Another vector system useful for delivering the polynucleotides of the present invention is the enterically administered recombinant poxvirus vaccines described by Small, Jr., P.A., et al. (U.S. Patent No. 5,676,950, issued October 14, 1997).

Additional viral vectors which will find use for delivering the nucleic acid molecules encoding the antigens of interest include those derived from the pox family of viruses, including vaccinia virus and avian poxvirus. By way of example, vaccinia virus recombinants expressing the genes can be constructed as follows. The DNA encoding the particular synthetic Gag/antigen coding sequence is first inserted into an appropriate vector so that it is adjacent to a vaccinia promoter and flanking vaccinia DNA sequences, such as the sequence encoding thymidine kinase (TK). This vector is then used to transfect cells which are simultaneously infected with vaccinia. Homologous recombination serves to insert the vaccinia promoter plus the gene encoding the coding sequences of interest into the viral genome. The resulting TK<sup>-</sup>recombinant can be selected by culturing the

cells in the presence of 5-bromodeoxyuridine and picking viral plaques resistant thereto.

Alternatively, avipoxviruses, such as the fowlpox and canarypox viruses, can also be used to deliver the genes. Recombinant avipox viruses, expressing immunogens from mammalian pathogens, are known to confer protective immunity when administered to non-avian species. The use of an avipox vector is particularly desirable in human and other mammalian species since members of the avipox genus can only productively replicate in susceptible avian species and therefore are not infective in mammalian cells. Methods for producing recombinant avipoxviruses are known in the art and employ genetic recombination, as described above with respect to the production of vaccinia viruses. See, e.g., WO 91/12882; WO 89/03429; and WO 92/03545.

Molecular conjugate vectors, such as the adenovirus chimeric vectors described in Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103, can also be used for gene delivery.

Members of the Alphavirus genus, such as, but not limited to, vectors derived from the Sindbis, Semliki Forest, and Venezuelan Equine Encephalitis viruses, will also find use as viral vectors for delivering the polynucleotides of the present invention (for example, a synthetic Gag- or Env-polypeptide encoding expression cassette as described in Example 14 below). For a description of Sindbis-virus derived vectors useful for the practice of the instant methods, see, Dubensky et al., *J. Virol.* (1996) 70:508-519; and International Publication Nos. WO 95/07995 and WO 96/17072; as well as, Dubensky, Jr., T.W., et al., U.S. Patent No. 5,843,723,



issued December 1, 1998, and Dubensky, Jr., T.W., U.S. Patent No. 5,789,245, issued August 4, 1998.

5 A vaccinia based infection/transfection system can be conveniently used to provide for inducible, transient expression of the coding sequences of interest (for example, a synthetic Gag/HCV-core expression cassette) in a host cell. In this system, cells are first infected in vitro with a vaccinia virus recombinant that encodes the bacteriophage T7 RNA polymerase. This polymerase  
10 displays exquisite specificity in that it only transcribes templates bearing T7 promoters. Following infection, cells are transfected with the polynucleotide of interest, driven by a T7 promoter. The polymerase expressed in the cytoplasm from the vaccinia virus  
15 recombinant transcribes the transfected DNA into RNA which is then translated into protein by the host translational machinery. The method provides for high level, transient, cytoplasmic production of large quantities of RNA and its translation products. See,  
20 e.g., Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA* (1990) 87:6743-6747; Fuerst et al., *Proc. Natl. Acad. Sci. USA* (1986) 83:8122-8126.

As an alternative approach to infection with vaccinia or avipox virus recombinants, or to the delivery  
25 of genes using other viral vectors, an amplification system can be used that will lead to high level expression following introduction into host cells. Specifically, a T7 RNA polymerase promoter preceding the coding region for T7 RNA polymerase can be engineered.  
30 Translation of RNA derived from this template will generate T7 RNA polymerase which in turn will transcribe more template. Concomitantly, there will be a cDNA whose expression is under the control of the T7 promoter. Thus, some of the T7 RNA polymerase generated from

translation of the amplification template RNA will lead to transcription of the desired gene. Because some T7 RNA polymerase is required to initiate the amplification, T7 RNA polymerase can be introduced into cells along with the template(s) to prime the transcription reaction. The polymerase can be introduced as a protein or on a plasmid encoding the RNA polymerase. For a further discussion of T7 systems and their use for transforming cells, see, e.g., International Publication No. WO 94/26911; Studier and Moffatt, *J. Mol. Biol.* (1986) 189:113-130; Deng and Wolff, *Gene* (1994) 143:245-249; Gao et al., *Biochem. Biophys. Res. Commun.* (1994) 200:1201-1206; Gao and Huang, *Nuc. Acids Res.* (1993) 21:2867-2872; Chen et al., *Nuc. Acids Res.* (1994) 22:2114-2120; and U.S. Patent No. 5,135,855.

The synthetic expression cassette of interest can also be delivered without a viral vector. For example, the synthetic expression cassette can be packaged as DNA or RNA in liposomes prior to delivery to the subject or to cells derived therefrom. Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed DNA to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight, *Biochim. Biophys. Acta.* (1991) 1097:1-17; Straubinger et al., in *Methods of Enzymology* (1983), Vol. 101, pp. 512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations, with cationic liposomes particularly preferred. Cationic liposomes have been shown to mediate intracellular

delivery of plasmid DNA (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416); mRNA (Malone et al., *Proc. Natl. Acad. Sci. USA* (1989) 86:6077-6081); and purified transcription factors (Debs et al., *J. Biol. Chem.* (1990) 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-7416). Other commercially available lipids include (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g., Szoka et al., *Proc. Natl. Acad. Sci. USA* (1978) 75:4194-4198; PCT Publication No. WO 90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as, from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See, e.g., Straubinger et al., in *METHODS OF*

IMMUNOLOGY (1983), Vol. 101, pp. 512-527; Szoka et al.,  
Proc. Natl. Acad. Sci. USA (1978) 75:4194-4198;  
Papahadjopoulos et al., Biochim. Biophys. Acta (1975)  
394:483; Wilson et al., Cell (1979) 17:77; Deamer and  
5 Bangham, Biochim. Biophys. Acta (1976) 443:629; Ostro et  
al., Biochem. Biophys. Res. Commun. (1977) 76:836; Fraley  
et al., Proc. Natl. Acad. Sci. USA (1979) 76:3348; Enoch  
and Strittmatter, Proc. Natl. Acad. Sci. USA (1979)  
76:145; Fraley et al., J. Biol. Chem. (1980) 255:10431;  
10 Szoka and Papahadjopoulos, Proc. Natl. Acad. Sci. USA  
(1978) 75:145; and Schaefer-Ridder et al., Science (1982)  
215:166.

The DNA and/or protein antigen(s) can also be  
delivered in cochleate lipid compositions similar to  
15 those described by Papahadjopoulos et al., Biochem.  
Biophys. Acta. (1975) 394:483-491. See, also, U.S.  
Patent Nos. 4,663,161 and 4,871,488.

The synthetic expression cassette of interest (e.g.,  
any of the synthetic expression cassettes described in  
20 Example 1) may also be encapsulated, adsorbed to, or  
associated with, particulate carriers. Such carriers  
present multiple copies of a selected antigen to the  
immune system and promote migration, trapping and  
retention of antigens in local lymph nodes. The  
25 particles can be taken up by professional antigen  
presenting cells such as macrophages and dendritic cells,  
and/or can enhance antigen presentation through other  
mechanisms such as stimulation of cytokine release.  
Examples of particulate carriers include those derived  
30 from polymethyl methacrylate polymers, as well as  
microparticles derived from poly(lactides) and  
poly(lactide-co-glycolides), known as PLG. See, e.g.,  
Jeffery et al., Pharm. Res. (1993) 10:362-368; McGee JP,

et al., *J Microencapsul.* 14(2):197-210, 1997; O'Hagan DT, et al., *Vaccine* 11(2):149-54, 1993.

Furthermore, other particulate systems and polymers can be used for the *in vivo* or *ex vivo* delivery of the gene of interest. For example, polymers such as polylysine, polyarginine, polyornithine, spermine, spermidine, as well as conjugates of these molecules, are useful for transferring a nucleic acid of interest.

Similarly, DEAE dextran-mediated transfection, calcium phosphate precipitation or precipitation using other insoluble inorganic salts, such as strontium phosphate, aluminum silicates including bentonite and kaolin, chromic oxide, magnesium silicate, talc, and the like, will find use with the present methods. See, e.g.,

Felgner, P.L., *Advanced Drug Delivery Reviews* (1990) 5:163-187, for a review of delivery systems useful for gene transfer. Peptoids (Zuckerman, R.N., et al., U.S. Patent No. 5,831,005, issued November 3, 1998) may also be used for delivery of a construct of the present invention.

Additionally, biolistic delivery systems employing particulate carriers such as gold and tungsten, are especially useful for delivering synthetic expression cassettes of the present invention. The particles are coated with the synthetic expression cassette(s) to be delivered and accelerated to high velocity, generally under a reduced atmosphere, using a gun powder discharge from a "gene gun." For a description of such techniques, and apparatuses useful therefore, see, e.g., U.S. Patent Nos. 4,945,050; 5,036,006; 5,100,792; 5,179,022; 5,371,015; and 5,478,744. Also, needle-less injection systems can be used (Davis, H.L., et al, *Vaccine* 12:1503-1509, 1994; Bioject, Inc., Portland, OR).

Recombinant vectors carrying a synthetic expression cassette of the present invention are formulated into compositions for delivery to the vertebrate subject. These compositions may either be prophylactic (to prevent infection) or therapeutic (to treat disease after infection). The compositions will comprise a "therapeutically effective amount" of the gene of interest such that an amount of the antigen can be produced *in vivo* so that an immune response is generated in the individual to which it is administered. The exact amount necessary will vary depending on the subject being treated; the age and general condition of the subject to be treated; the capacity of the subject's immune system to synthesize antibodies; the degree of protection desired; the severity of the condition being treated; the particular antigen selected and its mode of administration, among other factors. An appropriate effective amount can be readily determined by one of skill in the art. Thus, a "therapeutically effective amount" will fall in a relatively broad range that can be determined through routine trials.

The compositions will generally include one or more "pharmaceutically acceptable excipients or vehicles" such as water, saline, glycerol, polyethyleneglycol, hyaluronic acid, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, surfactants and the like, may be present in such vehicles. Certain facilitators of immunogenicity or of nucleic acid uptake and/or expression can also be included in the compositions or coadministered, such as, but not limited to, bupivacaine, cardiotoxin and sucrose.

Once formulated, the compositions of the invention can be administered directly to the subject (e.g., as

described above) or, alternatively, delivered ex vivo, to cells derived from the subject, using methods such as those described above. For example, methods for the ex vivo delivery and reimplantation of transformed cells into a subject are known in the art and can include, e.g., dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, lipofectamine and LT-1 mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) (with or without the corresponding antigen) in liposomes, and direct microinjection of the DNA into nuclei.

Direct delivery of synthetic expression cassette compositions in vivo will generally be accomplished with or without viral vectors, as described above, by injection using either a conventional syringe, needleless devices such as Bioject® or a gene gun, such as the Accell® gene delivery system (PowderJect Technologies, Inc., Oxford, England). The constructs can be delivered (e.g., injected) either subcutaneously, epidermally, intradermally, intramuscularly, intravenous, intramucosally (such as nasally, rectally and vaginally), intraperitoneally or orally. Delivery of DNA into cells of the epidermis is particularly preferred as this mode of administration provides access to skin-associated lymphoid cells and provides for a transient presence of DNA in the recipient. Other modes of administration include oral ingestion and pulmonary administration, suppositories, needle-less injection, transcutaneous and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule.

2.3.2      EX VIVO DELIVERY OF THE SYNTHETIC EXPRESSION  
CASSETTES OF THE PRESENT INVENTION

In one embodiment, T cells, and related cell types (including but not limited to antigen presenting cells, such as, macrophage, monocytes, lymphoid cells, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof), can be used for ex vivo delivery of the synthetic expression cassettes of the present invention. T cells can be isolated from peripheral blood lymphocytes (PBLs) by a variety of procedures known to those skilled in the art. For example, T cell populations can be "enriched" from a population of PBLs through the removal of accessory and B cells. In particular, T cell enrichment can be accomplished by the elimination of non-T cells using anti-MHC class II monoclonal antibodies. Similarly, other antibodies can be used to deplete specific populations of non-T cells. For example, anti-Ig antibody molecules can be used to deplete B cells and anti-MacI antibody molecules can be used to deplete macrophages.

T cells can be further fractionated into a number of different subpopulations by techniques known to those skilled in the art. Two major subpopulations can be isolated based on their differential expression of the cell surface markers CD4 and CD8. For example, following the enrichment of T cells as described above, CD4<sup>+</sup> cells can be enriched using antibodies specific for CD4 (see Coligan et al., *supra*). The antibodies may be coupled to a solid support such as magnetic beads. Conversely, CD8<sup>+</sup> cells can be enriched through the use of antibodies specific for CD4 (to remove CD4<sup>+</sup> cells), or can be isolated by the use of CD8 antibodies coupled to a solid support. CD4



lymphocytes from HIV-1 infected patients can be expanded ex vivo, before or after transduction as described by Wilson et. al. (1995) *J. Infect. Dis.* 172:88.

5        Following purification of T cells, a variety of methods of genetic modification known to those skilled in the art can be performed using non-viral or viral-based gene transfer vectors constructed as described herein. For example, one such approach involves transduction of  
10   the purified T cell population with vector-containing supernatant of cultures derived from vector producing cells. A second approach involves co-cultivation of an irradiated monolayer of vector-producing cells with the purified T cells. A third approach involves a similar  
15   co-cultivation approach; however, the purified T cells are pre-stimulated with various cytokines and cultured 48 hours prior to the co-cultivation with the irradiated vector producing cells. Pre-stimulation prior to such transduction increases effective gene transfer (Nolta et  
20   al. (1992) *Exp. Hematol.* 20:1065). Stimulation of these cultures to proliferate also provides increased cell populations for re-infusion into the patient. Subsequent to co-cultivation, T cells are collected from the vector producing cell monolayer, expanded, and frozen in liquid  
25   nitrogen.

Gene transfer vectors, containing one or more synthetic expression cassette of the present invention (associated with appropriate control elements for delivery to the isolated T cells) can be assembled using  
30   known methods.

Selectable markers can also be used in the construction of gene transfer vectors. For example, a marker can be used which imparts to a mammalian cell transduced with the gene transfer vector resistance to a

cytotoxic agent. The cytotoxic agent can be, but is not limited to, neomycin, aminoglycoside, tetracycline, chloramphenicol, sulfonamide, actinomycin, netropsin, distamycin A, anthracycline, or pyrazinamide. For example, neomycin phosphotransferase II imparts resistance to the neomycin analogue geneticin (G418).

The T cells can also be maintained in a medium containing at least one type of growth factor prior to being selected. A variety of growth factors are known in the art which sustain the growth of a particular cell type. Examples of such growth factors are cytokine mitogens such as rIL-2, IL-10, IL-12, and IL-15, which promote growth and activation of lymphocytes. Certain types of cells are stimulated by other growth factors such as hormones, including human chorionic gonadotropin (hCG) and human growth hormone. The selection of an appropriate growth factor for a particular cell population is readily accomplished by one of skill in the art.

For example, white blood cells such as differentiated progenitor and stem cells are stimulated by a variety of growth factors. More particularly, IL-3, IL-4, IL-5, IL-6, IL-9, GM-CSF, M-CSF, and G-CSF, produced by activated  $T_H$  and activated macrophages, stimulate myeloid stem cells, which then differentiate into pluripotent stem cells, granulocyte-monocyte progenitors, eosinophil progenitors, basophil progenitors, megakaryocytes, and erythroid progenitors. Differentiation is modulated by growth factors such as GM-CSF, IL-3, IL-6, IL-11, and EPO.

Pluripotent stem cells then differentiate into lymphoid stem cells, bone marrow stromal cells, T cell progenitors, B cell progenitors, thymocytes,  $T_H$  Cells,  $T_C$  cells, and B cells. This differentiation is modulated by

growth factors such as IL-3, IL-4, IL-6, IL-7, GM-CSF, M-CSF, G-CSF, IL-2, and IL-5.

Granulocyte-monocyte progenitors differentiate to monocytes, macrophages, and neutrophils. Such  
5 differentiation is modulated by the growth factors GM-CSF, M-CSF, and IL-8. Eosinophil progenitors differentiate into eosinophils. This process is modulated by GM-CSF and IL-5.

The differentiation of basophil progenitors into  
10 mast cells and basophils is modulated by GM-CSF, IL-4, and IL-9. Megakaryocytes produce platelets in response to GM-CSF, EPO, and IL-6. Erythroid progenitor cells differentiate into red blood cells in response to EPO.

Thus, during activation by the CD3-binding agent, T  
15 cells can also be contacted with a mitogen, for example a cytokine such as IL-2. In particularly preferred embodiments, the IL-2 is added to the population of T cells at a concentration of about 50 to 100  $\mu\text{g/ml}$ . Activation with the CD3-binding agent can be carried out  
20 for 2 to 4 days.

Once suitably activated, the T cells are genetically modified by contacting the same with a suitable gene transfer vector under conditions that allow for transfection of the vectors into the T cells. Genetic  
25 modification is carried out when the cell density of the T cell population is between about  $0.1 \times 10^6$  and  $5 \times 10^6$ , preferably between about  $0.5 \times 10^6$  and  $2 \times 10^6$ . A number of suitable viral and nonviral-based gene transfer vectors have been described for use herein.

30 After transduction, transduced cells are selected away from non-transduced cells using known techniques. For example, if the gene transfer vector used in the transduction includes a selectable marker which confers resistance to a cytotoxic agent, the cells can be

contacted with the appropriate cytotoxic agent, whereby non-transduced cells can be negatively selected away from the transduced cells. If the selectable marker is a cell surface marker, the cells can be contacted with a binding agent specific for the particular cell surface marker, whereby the transduced cells can be positively selected away from the population. The selection step can also entail fluorescence-activated cell sorting (FACS) techniques, such as where FACS is used to select cells from the population containing a particular surface marker, or the selection step can entail the use of magnetically responsive particles as retrievable supports for target cell capture and/or background removal.

More particularly, positive selection of the transduced cells can be performed using a FACS cell sorter (e.g. a FACSVantage™ Cell Sorter, Becton Dickinson Immunocytometry Systems, San Jose, CA) to sort and collect transduced cells expressing a selectable cell surface marker. Following transduction, the cells are stained with fluorescent-labeled antibody molecules directed against the particular cell surface marker. The amount of bound antibody on each cell can be measured by passing droplets containing the cells through the cell sorter. By imparting an electromagnetic charge to droplets containing the stained cells, the transduced cells can be separated from other cells. The positively selected cells are then harvested in sterile collection vessels. These cell sorting procedures are described in detail, for example, in the FACSVantage™ Training Manual, with particular reference to sections 3-11 to 3-28 and 10-1 to 10-17.

Positive selection of the transduced cells can also be performed using magnetic separation of cells based on expression of a particular cell surface marker. In such

separation techniques, cells to be positively selected are first contacted with specific binding agent (e.g., an antibody or reagent the interacts specifically with the cell surface marker). The cells are then contacted with  
5 retrievable particles (e.g., magnetically responsive particles) which are coupled with a reagent that binds the specific binding agent (that has bound to the positive cells). The cell-binding agent-particle complex can then be physically separated from non-labeled cells,  
10 for example using a magnetic field. When using magnetically responsive particles, the labeled cells can be retained in a container using a magnetic field while the negative cells are removed. These and similar separation procedures are known to those of ordinary  
15 skill in the art.

Expression of the vector in the selected transduced cells can be assessed by a number of assays known to those skilled in the art. For example, Western blot or Northern analysis can be employed depending on the nature  
20 of the inserted nucleotide sequence of interest. Once expression has been established and the transformed T cells have been tested for the presence of the selected synthetic expression cassette, they are ready for infusion into a patient via the peripheral blood stream.

25 The invention includes a kit for genetic modification of an ex vivo population of primary mammalian cells. The kit typically contains a gene transfer vector coding for at least one selectable marker and at least one synthetic expression cassette contained  
30 in one or more containers, ancillary reagents or hardware, and instructions for use of the kit.

**EXPERIMENTAL**

Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not  
5 intended to limit the scope of the present invention in any way.

Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should,  
10 of course, be allowed for.

Example 1Generation of Synthetic Gag and Env Expression Cassettes

15 A. Modification of HIV-1 Gag, Gag-protease, Gag-reverse transcriptase and Gag-polymerase Nucleic Acid Coding Sequences

The Gag (SEQ ID NO:1), Gag-protease (SEQ ID NO:2), Gag-polymerase (SEQ ID NO:3), and Gag-reverse  
20 transcriptase (SEQ ID NO:77) coding sequences were selected from the HIV-1SF2 strain (Sanchez-Pescador, R., et al., *Science* 227(4686): 484-492, 1985; Luciw, P.A., et al. U.S. Patent No. 5,156,949, issued October 20, 1992; Luciw, P.A., et al., U.S. Patent No. 5,688,688, November  
25 18, 1997). These sequences were manipulated to maximize expression of their gene products.

First, the HIV-1 codon usage pattern was modified so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human  
30 genes. The HIV codon usage reflects a high content of the nucleotides A or T of the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a high AU content in the RNA and in a decreased translation ability and instability of the

mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Gag-encoding sequences were modified to be comparable to codon usage found in highly expressed human genes.

5        Figure 11 presents a comparison of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN $\gamma$  mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH  
10 (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figure 11, the percent A-T content of these two sequences are compared to the percent A-T content of native HIV-1SF2 Gag cDNA and to the synthetic Gag cDNA  
15 sequence of the present invention. The top two panels of the figure show the percent A-T content over the length of the sequences for IFN $\gamma$  and native Gag. The bottom two panels of the figure show the percent A-T content over the length of the sequences for GAPDH and the synthetic  
20 Gag. Experiments performed in support of the present invention showed that the synthetic Gag sequences were capable of higher level of protein production (see the Examples) than the native Gag sequences. The data in Figure 11 suggest that one reason for this increased  
25 production may be increased stability of the mRNA corresponding to the synthetic Gag coding sequences versus the mRNA corresponding to the native Gag coding sequences.

30        Second, there are inhibitory (or instability) elements (INS) located within the coding sequences of the Gag and Gag-protease coding sequences (Schneider R, et al., *J Virol.* 71(7):4892-4903, 1997). RRE is a secondary RNA structure that interacts with the HIV encoded Rev-protein to overcome the expression down-regulating

effects of the INS. To overcome the requirement for post-transcriptional activating mechanisms of RRE and Rev, and to enhance independent expression of the Gag polypeptide, the INS were inactivated by introducing multiple point mutations that did not alter the reading frame of the encoded proteins. Figure 1 shows the original SF2 Gag sequence, the location of the INS sequences, and the modifications made to the INS sequences to reduce their effects.

For the Gag-protease sequence (wild type, SEQ ID NO:2; synthetic, SEQ ID NOs:5, 78 and 79), the changes in codon usage were restricted to the regions up to the -1 frameshift and starting again at the end of the Gag reading frame (Figure 2; the region indicated in lower case letters in Figure 2 is the unmodified region). Further, inhibitory (or instability) elements (INS) located within the coding sequences of the Gag-protease polypeptide coding sequence were altered as well (indicated in Figure 2). The synthetic coding sequences were assembled by the Midland Certified Reagent Company (Midland, Texas).

Modification of the Gag-polymerase sequences (wild type, SEQ ID NO:3; synthetic, SEQ ID NO:6) and Gag-reverse transcriptase sequences (SEQ ID NOs:80 through 84) include similar modifications as described for Gag-protease in order to preserve the frameshift region. Locations of the inactivation sites and changes to the sequence to alter the inactivation sites are presented in Figure 12 for the native HIV-1<sub>SF2</sub> Gag-polymerase sequence.

In one embodiment of the invention, the full length polymerase coding region of the Gag-polymerase sequence is included with the synthetic Gag sequences in order to increase the number of epitopes for virus-like particles expressed by the synthetic, optimized Gag expression



cassette. Because synthetic HIV-1 Gag-polymerase expresses the potentially deleterious functional enzymes reverse transcriptase (RT) and integrase (INT) (in addition to the structural proteins and protease), it is important to inactivate RT and INT functions. Several in-frame deletions in the RT and INT reading frame can be made to achieve catalytic nonfunctional enzymes with respect to their RT and INT activity. {Jay. A. Levy (Editor) (1995) *The Retroviridae*, Plenum Press, New York. ISBN 0-306-45033X. Pages 215-20; Grimison, B. and Laurence, J. (1995), *Journal Of Acquired Immune Deficiency Syndromes and Human Retrovirology* 9(1):58-68; Wakefield, J. K., et al., (1992) *Journal Of Virology* 66(11):6806-6812; Esnouf, R., et al., (1995) *Nature Structural Biology* 2(4):303-308; Maignan, S., et al., (1998) *Journal Of Molecular Biology* 282(2):359-368; Katz, R. A. and Skalka, A. M. (1994) *Annual Review Of Biochemistry* 73 (1994); Jacobo-Molina, A., et al., (1993) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 90(13):6320-6324; Hickman, A. B., et al., (1994) *Journal Of Biological Chemistry* 269(46):29279-29287; Goldgur, Y., et al., (1998) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 95(16):9150-9154; Goette, M., et al., (1998) *Journal Of Biological Chemistry* 273(17):10139-10146; Gorton, J. L., et al., (1998) *Journal of Virology* 72(6):5046-5055; Engelman, A., et al., (1997) *Journal Of Virology* 71(5):3507-3514; Dyda, F., et al., *Science* 266(5193):1981-1986; Davies, J. F., et al., (1991) *Science* 252(5002):88-95; Bujacz, G., et al., (1996) *Febs Letters* 398(2-3):175-178; Beard, W. A., et al., (1996) *Journal Of Biological Chemistry* 271(21):12213-12220; Kohlstaedt, L. A., et al., (1992)

Science 256(5065):1783-1790; Krug, M. S. and Berger, S. L. (1991) *Biochemistry* 30(44):10614-10623; Mazumder, A., et al., (1996) *Molecular Pharmacology* 49(4):621-628; Palaniappan, C., et al., (1997) *Journal Of Biological Chemistry* 272(17):11157-11164; Rodgers, D. W., et al., (1995) *Proceedings Of the National Academy Of Sciences Of the United States Of America* 92(4):1222-1226; Sheng, N. and Dennis, D. (1993) *Biochemistry* 32(18):4938-4942; Spence, R. A., et al., (1995) *Science* 267(5200):988-993.

Furthermore selected B- and/or T-cell epitopes can be added to the Gag-polymerase constructs within the deletions of the RT- and INT-coding sequence to replace and augment any epitopes deleted by the functional modifications of RT and INT. Alternately, selected B- and T-cell epitopes (including CTL epitopes) from RT and INT can be included in a minimal VLP formed by expression of the synthetic Gag or synthetic GagProt cassette, described above. (For descriptions of known HIV B- and T-cell epitopes see, HIV Molecular Immunology Database CTL Search Interface; Los Alamos Sequence Compendia, 1987-1997; Internet address: <http://hiv-web.lanl.gov/immunology/index.html>.)

The resulting modified coding sequences are presented as a synthetic Gag expression cassette (SEQ ID NO:4), a synthetic Gag-protease expression cassette (SEQ ID NOs:5, 78 and 79), and a synthetic Gag-polymerase expression cassette (SEQ ID NO:6). Synthetic expression cassettes containing codon modifications in the reverse transcriptase region are shown in SEQ ID NOs:80 through 84. An alignment of selected sequences is presented in Figure 7. A common region (Gag-common; SEQ ID NO:9) extends from position 1 to position 1262.

The synthetic DNA fragments for Gag and Gag-protease were cloned into the following expression vectors:

pCMVKm2, for transient expression assays and DNA immunization studies, the pCMVKm2 vector was derived from pCMV6a (Chapman et al., *Nuc. Acids Res.* (1991) 19:3979-3986) and comprises a kanamycin selectable marker, a  
5 ColE1 origin of replication, a CMV promoter enhancer and Intron A, followed by an insertion site for the synthetic sequences described below followed by a polyadenylation signal derived from bovine growth hormone -- the pCMVKm2 vector differs from the pCMV-link vector only in that a  
10 polylinker site was inserted into pCMVKm2 to generate pCMV-link (Figure 14, polylinker at positions 1646 to 1697); pESN2dhfr (Figure 13A) and pCMVPLEdhfr (also known as pCMVIII as shown in Figure 13B), for expression in Chinese Hamster Ovary (CHO) cells; and, pAcc13, a shuttle  
15 vector for use in the Baculovirus expression system (pAcc13, was derived from pAcc12 which was described by Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990).

A restriction map for vector pCMV-link is presented  
20 in Figure 14. In the figure, the CMV promoter (CMV IE ENH/PRO), bovine growth hormone terminator (BGH pA), kanamycin selectable marker (kan), and a ColE1 origin of replication (ColE1 ori) are indicated. A polycloning site is also indicated in the figure following the CMV  
25 promoter sequences.

A restriction map for vector pESN2dhfr is presented in Figure 13A. In the figure, the CMV promoter (pCMV, hCMVIE), bovine growth hormone terminator (BGHpA), SV40 origin of replication (SV40ori), neomycin selectable  
30 marker (Neo), SV40 polyA (SV40pA), Adenovirus 2 late promoter (Ad2VLP), and the murine dhfr gene (mu dhfr) are indicated. A polycloning site is also indicated in the figure following the CMV promoter sequences.

Briefly, construction of pCMVPLEdhfr (pCMVIII) was as follows. To construct a DHFR cassette, the EMCV IRES (internal ribosome entry site) leader was PCR-amplified from pCite-4a+ (Novagen, Inc., Milwaukee, WI) and  
5 inserted into pET-23d (Novagen, Inc., Milwaukee, WI) as an Xba-Nco fragment to give pET-EMCV. The *dhfr* gene was PCR-amplified from pESN2dhfr to give a product with a Gly-Gly-Gly-Ser spacer in place of the translation stop codon and inserted as an Nco-BamH1 fragment to give pET-  
10 E-DHFR. Next, the attenuated *neo* gene was PCR amplified from a pSV2Neo (Clontech, Palo Alto, CA) derivative and inserted into the unique BamH1 site of pET-E-DHFR to give pET-E-DHFR/Neo<sub>(m2)</sub>. Then, the bovine growth hormone terminator from pCDNA3 (Invitrogen, Inc., Carlsbad, CA)  
15 was inserted downstream of the *neo* gene to give pET-E-DHFR/Neo<sub>(m2)</sub>BGHT. The EMCV-*dhfr*/*neo* selectable marker cassette fragment was prepared by cleavage of pET-E-DHFR/Neo<sub>(m2)</sub>BGHT. The CMV enhancer/promoter plus Intron A was transferred from pCMV6a (Chapman et al., *Nuc. Acids*  
20 *Res.* (1991) 19:3979-3986) as a HindIII-SalI fragment into pUC19 (New England Biolabs, Inc., Beverly, MA). The vector backbone of pUC19 was deleted from the NdeI to the SapI sites. The above described DHFR cassette was added to the construct such that the EMCV IRES followed the CMV  
25 promoter to produce the final construct. The vector also contained an *amp<sup>r</sup>* gene and an SV40 origin of replication.

Selected pCMVKm2 vectors containing the synthetic expression cassettes have been designated as follows: pCMVKm2.GagMod.SF2, pCMVKm2.GagprotMod.SF2, and  
30 pCMVKm2.GagpolMod.SF2, pCMVKm2.GagprotMod.SF2.GP1 (SEQ ID NO:78) and pCMVKm2.GagprotMod.SF2.GP2 (SEQ ID NO:79). Other exemplary Gag-encoding expressing cassettes are shown in the Figures and as Sequence Listings.

B. Modification of HIV-1 Gag/Hepatitis C Core Chimeric Protein Nucleic Acid Coding Sequences Generation of Synthetic Expression Cassettes

To facilitate the ligation of the Gag and HCV core coding sequences, PCR amplification was employed. The synthetic p55Gag expression cassette was used as a PCR template with the following primers: GAG5 (SEQ ID NO:11) and P55-SAL3 (SEQ ID NO:12). The PCR amplification was conducted at 55°C for 25 cycles using Stratagene's Pfu polymerase. The resulting PCR product was rendered free of nucleotides and primers using the Promega PCR clean-up kit and then subjected to EcoRI and SalI digestions. For HCV core coding sequences, the following primers were used with an HCV template (Houghton, M., et al., U.S. Patent No. 5,714,596, issued February 3, 1998; Houghton, M., et al., U.S. Patent No. 5,712,088, issued January 27, 1998; Houghton, M., et al., U.S. Patent No. 5,683,864, issued November 4, 1997; Weiner, A.J., et al., U.S. Patent No. 5,728,520, issued March 17, 1998; Weiner, A.J., et al., U.S. Patent No. 5,766,845, issued June 16, 1998; Weiner, A.J., et al., U.S. Patent No. 5,670,152, issued September 23, 1997): CORESAL 5 (SEQ ID NO:13) and 173CORE (SEQ ID NO:14) using the conditions outlined above. The purified product was digested with SalI and BamHI restriction enzymes. The digested Gag and HCV core PCR products were ligated into the pCMVKm2 vector digested with EcoRI and BamHI. Ligation of the PCR products at the SalI site resulted in a direct fusion of the final amino acid of p55Gag to the second amino acid of HCV core, serine. Amino acid 173 of core is a serine and is followed immediately by a TAG termination codon. The sequence of the fusion clone was confirmed. The pCMVKm2 vector containing the synthetic expression

cassette was designated as pCMVKm2.GagModHCVcore.

The EcoRI-BamHI fragment of p55Gag-core 173 was also cloned into EcoRI-BamHI-digested pAcC13 for baculovirus expression. Western blots confirmed expression and sucrose gradient sedimentation along with electron microscopy confirmed particle formation. To generate the above clone but containing the synthetic Gag sequences (instead of wild-type), the following steps were performed: pCMVKm2-modified p55Gag was used as template for PCR amplification with MS65 (SEQ ID NO:15) and MS66 (SEQ ID NO:16) primers. The region amplified corresponds to the BspHI and SalI sites at the C-terminus of synthetic Gag sequence. The amplification product was digested with BspHI and SalI and ligated to SalI/BamHI digested pCMV-link along with the Sal/BspHI fragment from pCMV-Km-p55modGag, representing the amino terminal end of modified Gag, and the SalI/BamHI fragment from pCMV-p55Gag-core173. Thereafter, a T4-blunted-SalI partial/BamHI fragment was ligated into pAcC4-SmaI/BamHI to generate pAcC4-p55GagMod-core173 (containing the synthetic sequence presented as SEQ ID NO:7).

C. Defining of the Major Homology Region (MHR) of HIV-1 p55Gag

The Major Homology Region (MHR) of HIV-1 p55 (Gag) is located in the p24-CA sequence of Gag. It is a conserved stretch of 20 amino acids (SEQ ID NO:19). The position in the wild type HIV-1<sub>SF2</sub> Gag protein is from aa 286-305 and spans a region from nucleotides 856-915 in the native HIV-1<sub>SF2</sub> Gag DNA-sequence. The position in the synthetic Gag protein is from aa 288-307 and spans a region from nucleotides 862-921 for the synthetic Gag DNA-sequence. The nucleotide sequence for the MHR in the synthetic

GagMod.SF2 is presented as SEQ ID NO:20. Mutations or deletions in the amino acid sequence of the MHR can severely impair particle production (Borsetti, A., et al., *J. Virol.* 72(11):9313-9317, 1998; Mammano, F., et al., *J Virol* 68(8):4927-4936, 1994).

Percent identity to the MHR nucleotide sequence can be determined, for example, using the MacDNAsis program (Hitachi Software Engineering America Limited, South San Francisco, CA), Higgins algorithm, with the following exemplary parameters: gap penalty = 5, no. of top diagonals = 5, fixed gap penalty = 5, K-tuple = 2, window size = 5, and floating gap penalty = 10.

#### D. Generation of Synthetic Env Expression Cassettes

Env coding sequences of the present invention include, but are not limited to, polynucleotide sequences encoding the following HIV-encoded polypeptides: gp160, gp140, and gp120 (see, e.g., U.S. Patent No. 5,792,459 for a description of the HIV-1<sub>SF2</sub> ("SF2") Env polypeptide). The relationships between these polypeptides is shown schematically in Figure 15 (in the figure: the polypeptides are indicated as lines, the amino and carboxy termini are indicated on the gp160 line; the open circle represents the oligomerization domain; the open square represents a transmembrane spanning domain (TM); and "c" represents the location of a cleavage site, in gp140.mut the "X" indicates that the cleavage site has been mutated such that it no longer functions as a cleavage site). The polypeptide gp160 includes the coding sequences for gp120 and gp41. The polypeptide gp41 is comprised of several domains including an oligomerization domain (OD) and a transmembrane spanning domain (TM). In the native envelope, the oligomerization domain is required for the

non-covalent association of three gp41 polypeptides to form a trimeric structure: through non-covalent interactions with the gp41 trimer (and itself), the gp120 polypeptides are also organized in a trimeric structure.

5 A cleavage site (or cleavage sites) exists approximately between the polypeptide sequences for gp120 and the polypeptide sequences corresponding to gp41. This cleavage site(s) can be mutated to prevent cleavage at the site. The resulting gp140 polypeptide corresponds to

10 a truncated form of gp160 where the transmembrane spanning domain of gp41 has been deleted. This gp140 polypeptide can exist in both monomeric and oligomeric (i.e. trimeric) forms by virtue of the presence of the oligomerization domain in the gp41 moiety. In the

15 situation where the cleavage site has been mutated to prevent cleavage and the transmembrane portion of gp41 has been deleted the resulting polypeptide product is designated "mutated" gp140 (e.g., gp140.mut). As will be apparent to those in the field, the cleavage site can be

20 mutated in a variety of ways. The native amino acid sequence in the SF162 cleavage sites is: APTKAKRRVVQREKR (SEQ ID NO:21), where KAKRR (SEQ ID NO:22) is termed the "second" site and REKR (SEQ ID NO:23) is the "first site". Exemplary mutations include the following

25 constructs: gp140.mut7.modSF162 which encodes the amino acid sequence APTKAISSVVQSEKS (SEQ ID NO:24) in the cleavage site region; gp140.mut8.modSF162 which encodes the amino acid sequence APTIAISSVVQSEKS (SEQ ID NO:25) in the cleavage site region and gp140mut.modSF162 which  
30 encodes the amino acid sequence APTKAKRRVVQREKS (SEQ ID NO:26). Mutations are denoted in bold. The native amino acid sequence in the US4 cleavage sites is:

APTQAKRRVVQREKR (SEQ ID NO:27), where QAKRR (SEQ ID NO:28) is termed the "second" site and REKR (SEQ ID



NO:23) is the "first site". Exemplary mutations include the following construct: gp140.mut.modUS4 which encodes the amino acid sequence APTQAKRRVVQREKS (SEQ ID NO:29) in the cleavage site region. Mutations are denoted in bold.

5

E. Modification of HIV-1 Env (Envelope) Nucleic Acid Coding Sequences

In one embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-1<sub>SF162</sub> ("SF162") strain (Cheng-Mayer (1989) *PNAS USA* 86:8575-8579). These SF162 sequences were as follows: gp120, SEQ ID NO:30 (Fig. 16); gp140, SEQ ID NO:31 (Fig. 17); and gp160, SEQ ID NO:32 (Fig. 18).

10 In another embodiment of the present invention, wild-type Env coding sequences were selected from the HIV-US4 strain (Mascola, et al. (1994) *J. Infect. Dis.* 169:48-54). These US4 sequences were as follows: gp120, SEQ ID NO:51 (Fig. 38); gp140, SEQ ID NO:52 (Fig. 39); and gp160, SEQ ID NO:53 (Fig. 40).

15 These Env coding sequences were manipulated to maximize expression of their gene products.

First, the wild-type coding region was modified in one or more of the following ways. In one embodiment, sequences encoding hypervariable regions of Env, particularly V1 and/or V2 were deleted. In other embodiments, mutations were introduced into sequences encoding the cleavage site in Env to abrogate the enzymatic cleavage of oligomeric gp140 into gp120 monomers. (See, e.g., Earl et al. (1990) *PNAS USA* 87:648-652; Earl et al. (1991) *J. Virol.* 65:31-41). In yet other embodiments, hypervariable region(s) were deleted, N-glycosylation sites were removed and/or cleavage sites mutated.

20 Second, the HIV-1 codon usage pattern was modified

so that the resulting nucleic acid coding sequence was comparable to codon usage found in highly expressed human genes. The HIV codon usage reflects a high content of the nucleotides A or T in the codon-triplet. The effect of the HIV-1 codon usage is a high AT content in the DNA sequence that results in a decreased translation ability and instability of the mRNA. In comparison, highly expressed human codons prefer the nucleotides G or C. The Env coding sequences were modified to be comparable to codon usage found in highly expressed human genes.

Figures 22A-22H present comparisons of the percent A-T content for the cDNAs of stable versus unstable RNAs (comparison window size = 50). Human IFN $\gamma$  mRNA is known to (i) be unstable, (ii) have a short half-life, and (iii) have a high A-U content. Human GAPDH (glyceraldehyde-3-phosphate dehydrogenase) mRNA is known to (i) be a stable RNA, and (i) have a low A-U content. In Figures 22A-H, the percent A-T content of these two sequences are compared to the percent A-T content of (1) native HIV-1 US4 Env gp160 cDNA, a synthetic US4 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention; and (2) native HIV-1 SF162 Env gp160 cDNA, a synthetic SF162 Env gp160 cDNA sequence (i.e., having modified codons) of the present invention. Figures 22A-H show the percent A-T content over the length of the sequences for IFN $\gamma$  (Figures 22C and 22G); native gp160 Env US4 and SF162 (Figures 22A and 22E, respectively); GAPDH (Figures 22D and 22H); and the synthetic gp160 Env for US4 and SF162 (Figures 22B and 22F). Experiments performed in support of the present invention showed that the synthetic Env sequences were capable of higher level of protein production (see the Examples) than the native Env sequences. The data in Figures 22A-H suggest that one reason for this increased

production is increased stability of the mRNA corresponding to the synthetic Env coding sequences versus the mRNA corresponding to the native Env coding sequences.

5           To create the synthetic coding sequences of the present invention the gene cassettes were designed to comprise the entire coding sequence of interest. Synthetic gene cassettes were constructed by oligonucleotide synthesis and PCR amplification to  
10 generate gene fragments. Primers were chosen to provide convenient restriction sites for subcloning. The resulting fragments were then ligated to create the entire desired sequence which was then cloned into an appropriate vector. The final synthetic sequences were  
15 (i) screened by restriction endonuclease digestion and analysis, (ii) subjected to DNA sequencing in order to confirm that the desired sequence had been obtained and (iii) the identity and integrity of the expressed protein confirmed by SDS-PAGE and Western blotting (See,  
20 Examples. The synthetic coding sequences were assembled at Chiron Corp. or by the Midland Certified Reagent Company (Midland, Texas).

Exemplary modified coding sequences are presented as synthetic Env expression cassettes in Table 1A and 1B.  
25 The following expression cassettes (i) have unique, terminal *EcoRI* and *XbaI* cloning sites; (ii) include Kozak sequences to promote optimal translation; (iii) tPA signal sequences (to direct the ENV polypeptide to the cell membrane, see, e.g., Chapman et al., *infra*); (iv)  
30 open reading frames optimized for expression in mammalian cells; and (v) a translational stop signal codon.

Table 1A: Exemplary Synthetic Env Expression  
Cassettes(SF162)

	Expression Cassette	Seq Id	Further Information
5	gp120 SF162	30	wild-type; Figure 16
	gp140 SF162	31	wild-type; Figure 17
	gp160 SF162	32	wild-type; Figure 18
	gp120.modSF162	33	none; Figure 19
	gp120.modSF162.delV2	34	deleted V2 loop; Figure 20
10	gp120.modSF162.delV1/V2	35	deleted V1 and V2; Figure 21
	gp140.modSF162	36	none; Figure 23
	gp140.modSF162.delV2	37	deleted V2 loop; Figure 24
	gp140.modSF162.delV1/V2	38	deleted V1 and V2; Figure 25
	gp140.mut.modSF162	39	mutated cleavage site; Fig. 26
15	gp140.mut.modSF162.delV2	40	deleted V2; mutated cleavage site; Figure 27
	gp140.mut.modSF162.delV1/V 2	41	deleted V1 & V2; mutated cleavage site; Figure 28
	gp140.mut7.modSF162	42	mutated cleavage site; Fig. 29
	gp140.mut7.modSF162.delV2	43	mutated cleavage site; deleted V2; Figure 30
20	gp140.mut7.modSF162.delV1/ V2	44	mutated cleavage site; deleted V1 and V2; Figure 31
	gp140.mut8.modSF162	45	mutated cleavage site; Fig. 32
	gp140.mut8.modSF162.delV2	46	mutated cleavage site; deleted V2; Figure 33
25	gp140.mut8.modSF162.delV1/ V2	47	mutated cleavage site; deleted V1 and V2; Figure 34
	gp160.modSF162	48	none; Figure 35
	gp160.modSF162.delV2	49	deleted V2 loop; Figure 36
	gp160.modSF162.delV1/V2	50	deleted V1 & V2; Figure 37

Table 1B:

## Exemplary Synthetic Env Expression Cassettes (US4)

Expression Cassette	Seq Id	Further Information
gp120 US4	51	wild-type; Figure 38
gp140 US4	52	wild-type; Figure 39
gp160 US4	53	wild-type; Figure 40
gp120.modUS4	54	none; Figure 41
gp120.modUS4.del 128-194	55	deletion in V1 and V2 regions; Figure 42
gp140.modUS4	56	none; Figure 43
gp140.modUS4	57	mutated cleavage site; Figure 44
gp140TM.modUS4	58	native transmembrane region; Figure 45
gp140.modUS4.delV1/V2	59	deleted V1 and V2; Figure 46
gp140.modUS4.delV2	60	deleted V1; Figure 47
gp140.modUS4.delV1/V2	61	mutated cleavage site; deleted V1 and V2; Figure 48
gp140.modUS4.del 128-194	62	deletion in V1 and V2 regions; Figure 49
gp140.modUS4.del 128-194	63	mutated cleavage site; deletion in V1 and V2 regions; Figure 50
gp160.modUS4	64	none; Figure 51
gp160.modUS4.delV1	65	deleted V1; Figure 52
gp160.modUS4.delV2	66	deleted V2; Figure 53
gp160.modUS4.delV1/V2	67	deleted V1 and V2; Figure 54
gp160.modUS4del 128-194	68	deletion in V1 and V2 regions; Figure 55

Alignments of the sequences presented in the above  
tables are presented in Figures 66A and 66B.

A common region (Env-common) extends from nucleotide  
position 1186 to nucleotide position 1329 (SEQ ID NO:69,

Fig. 56) relative to the wild-type US4 sequence and from nucleotide position 1117 to position 1260 (SEQ ID NO:79, Fig. 57) relative to the wild-type SF162 sequence. The synthetic sequences of the present invention

5 corresponding to these regions are presented, as SEQ ID NO:71 (Figure 58) for the synthetic Env US4 common region and as SEQ ID NO:72 (Figure 59) for the synthetic Env SF162 common region.

Percent identity to this sequence can be determined, 10 for example, using the Smith-Waterman search algorithm (Time Logic, Incline Village, NV), with the following exemplary parameters: weight matrix = nuc4x4hb; gap opening penalty = 20, gap extension penalty = 5, reporting threshold = 1; alignment threshold = 20.

15 Various forms of the different embodiments of the present invention (e.g., constructs) may be combined.

F. Cloning Synthetic Env Expression Cassettes of the Present Invention.

20 The synthetic DNA fragments encoding the Env polypeptides were typically cloned into the eucaryotic expression vectors described above for Gag, for example, pCMVKm2/pCMVlink (Figure 4), pCMV6a, pESN2dhfr (Figure 13A), pCMVIII (Figure 13B; alternately designated as the 25 pCMV-PL-E-dhfr/neo vector).

Exemplary designations for pCMVlink vectors containing synthetic expression cassettes of the present invention are as follows: pCMVlink.gp140.modSF162; pCMVlink.gp140.-modSF162.delV2; 30 pCMVlink.gp140.mut.modSF162; pCMVlink.gp140.mut.modSF162.delV2; pCMVKm2.gp140modUS4; pCMVKm2.gp140.modUS4.delV2; pCMVKm2.gp140.mut.modUS4; and, pCMVKm2.gp140.mut.modUS4.delV1/V2.

G. Generation of Synthetic Tat Expression Cassettes

Tat coding sequences have also been modified according to the teachings of the present specification. The wild type nucleotide sequence encoding tat from variant SF162 is presented in Figure 76 (SEQ ID NO:85). The corresponding wild-type amino acid sequence is presented in Figure 77 (SEQ ID NO:86). Figure 81 (SEQ ID NO:89) shows the nucleotide sequence encoding the amino terminal of the tat protein and the codon encoding cystein-22 is underlined. Other exemplary constructs encoding synthetic tat polypeptides are shown in Figures 78 and 79 (SEQ ID NOs:87 and 88). In one embodiment (SEQ ID NO:88), the cystein residue at position 22 is replaced by a glycine. Caputo et al. (1996) *Gene Therapy* 3:235 have shown that this mutation affects the trans activation domain of Tat.

Various forms of the different embodiments of the invention, described herein, may be combined.

H. Deposit of Vectors

Selected exemplary constructs shown below and described herein are deposited at Chiron Corporation, Emeryville, CA, 94662-8097, and were sent to the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209 on December 27, 1999.

	Plasmid Name	Chiron Deposit #	Date Sent to ATCC
	pCMVgpl60.modUS4	5094	27 Dec 99
	pCMVgpl60delI.modUS4	5095	27 Dec 99
	pCMVgpl60del2.modUS4	5096	27 Dec 99
5	pCMVgpl60del-2.modUS4	5097	27 Dec 99
	pCMVgpl60del128-194.mod.US4	5098	27 Dec 99
	pCMVgpl40mut.modUS4del128-194	5100	27 Dec 99
	pCMVgpl40.mut.mod.US	5101	27 Dec 99
	pCMVgpl60.modSF162	5125	27 Dec 99
10	pCMVgpl60.modSF162.delV2	5126	27 Dec 99
	pCMVgpl60.modSF162.delV1V2	5127	27 Dec 99
	pCMVgpl40.mut.modSF162delV2	5128	27 Dec 99
	pCMVgpl40.mut7.modSF162	5129	27 Dec 99
	pCMVgpl40.mut7.modSF162delV2	5130	27 Dec 99
15	pCMVgpl40.mut8.modSF162	5131	27 Dec 99
	pCMVgpl40.mut8.modSF162delV2	5132	27 Dec 99
	pCMVgpl40.mut8.modSF162delV1V2	5133	27 Dec 99
	pCMVKm2.Gagprot.Mod.SF2.GP1	5150	27 Dec 99
20	pCMVKm2.Gagprot.Mod.SF2.GP2	5151	27 Dec 99

### Example 2

#### Expression Assays for the

#### Synthetic Gag, Env and Tat Coding Sequences

#### 25 A. Gag and Gag-Protease Coding Sequences

The HIV-1SF2 wild-type Gag (SEQ ID NO:1) and Gag-protease (SEQ ID NO:2) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Gag (SEQ ID NO:4) and  
 30 Gag-protease (SEQ ID NOs:5, 78 or 79)) sequences were cloned.



Expression efficiencies for various vectors carrying the HIV-1SF2 wild-type and synthetic Gag sequences were evaluated as follows. Cells from several mammalian cell lines (293, RD, COS-7, and CHO; all obtained from the American Type Culture Collection, 10801 University Boulevard, Manassas, VA 20110-2209) were transfected with 2  $\mu$ g of DNA in transfection reagent LT1 (PanVera Corporation, 545 Science Dr., Madison, WI). The cells were incubated for 5 hours in reduced serum medium (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was then replaced with normal medium as follows: 293 cells, IMDM, 10% fetal calf serum, 2% glutamine (BioWhittaker, Walkersville, MD); RD and COS-7 cells, D-MEM, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD); and CHO cells, Ham's F-12, 10% fetal calf serum, 2% glutamine (Opti-MEM, Gibco-BRL, Gaithersburg, MD). The cells were incubated for either 48 or 60 hours. Supernatants were harvested and filtered through 0.45  $\mu$ m syringe filters and, optionally, stored at -20°C.

Supernatants were evaluated using the Coulter p24-assay (Coulter Corporation, Hialeah, FL, US), using 96-well plates coated with a murine monoclonal antibody directed against HIV core antigen. The HIV-1 p24 antigen binds to the coated wells. Biotinylated antibodies against HIV recognize the bound p24 antigen. Conjugated streptavidin-horseradish peroxidase reacts with the biotin. Color develops from the reaction of peroxidase with TMB substrate. The reaction is terminated by addition of 4N H<sub>2</sub>SO<sub>4</sub>. The intensity of the color is directly proportional to the amount of HIV p24 antigen in a sample.

The results of these expression assays are presented in Tables 2A and 2B. Tables 2A and 2B shows data

obtained using the synthetic Gag-protease expression cassette of SEQ ID NO:5. Similar results were obtained using the Gag-protease expression cassettes of SEQ ID NOs:78 and 79.

5

Table 2: in vitro gag and gagprot p24 expression

5 TABLE 2a. Increased in vitro expression from modified vs. native gag plasmids in supernatants and lysates from transiently transfected cells

experiment	native (nat) <sup>a</sup> modified (mod) <sup>b</sup>	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 (fold increase)
1	nat	sup	293	48	3.4
	mod	sup	293	48	1260 (371)
	nat	sup	293	60	3.2
	mod	sup	293	60	2222 (694)
2	nat	sup	293	60	1.8
	mod	sup	293	60	1740 (966)
3	nat	sup	293	60	1.8
	mod	sup	293	60	580 (322)
4	nat	lys	293	60	1.5
	mod	lys	293	60	85 (57)
1	nat	sup	RD	48	5.6
	mod	sup	RD	48	66 (12)
	nat	sup	RD	60	7.8
	mod	sup	RD	60	70.2 (9)
2	nat	lys	RD	60	1.9
	mod	lys	RD	60	7.8 (4)
1	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	33.4 (84)
2	nat	sup	COS-7	48	0.4
	mod	sup	COS-7	48	10 (25)
	nat	lys	COS-7	48	3
	mod	lys	COS-7	48	14 (5)

<sup>a</sup> pCMVLink.Gag.SF2.PRE

<sup>b</sup> pCMVKm2.GagMod.SF2

5 TABLE 2b. *In vitro* expression from modified *gag* and *gagprotease* plasmids in supernatants and lysates from transiently transfected cells

plasmid	supernatant (sup) lysate (lys)	cell line	hours post transfection	total ng p24 <sup>d</sup>
Gag <sup>a</sup>	sup	293	60	760
GagProt(GP1) <sup>b</sup>	sup	293	60	380
GagProt(GP2) <sup>c</sup>	sup	293	60	320
Gag	lys	293	60	78
GagProt(GP1)	lys	293	60	1250
GagProt(GP2)	lys	293	60	400
Gag	sup	COS-7	72	40
GagProt(GP1)	sup	COS-7	72	150
GagProt(GP2)	sup	COS-7	72	290
Gag	lys	COS-7	72	60
GagProt(GP1)	lys	COS-7	72	63
GagProt(GP2)	lys	COS-7	72	58

<sup>a</sup> pCMVKm2.GagMod.SF2

<sup>b</sup> pCMVKm2.GagProtMod.SF2(GP1) *gagprotease* with codon optimization and inactivation of INS in *protease*

<sup>c</sup> pCMVKm2.GagProtMod.SF2(GP2) *gagprotease* with only inactivation of INS in *protease*

<sup>d</sup> Shown are representative results from 3 independent experiments for each cell line tested.

The data showed that the synthetic Gag and Gag-protease expression cassettes provided dramatic increases in production of their protein products, relative to the native (HIV-1SF2 wild-type) sequences, when expressed in a variety of cell lines.

#### B. Env Coding Sequences

The HIV-SF162 ("SF162") wild-type Env (SEQ ID NO:1-3) and HIV-US4 ("US4") wild-type Env (SEQ ID NO:22-24) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Env sequences were cloned.

Expression efficiencies for various vectors carrying the SF162 and US4 wild-type and synthetic Env sequences were evaluated essentially as described above for Gag except that cell lysates were prepared in 40  $\mu$ l lysis buffer (1.0 % NP40, 0.1 M Tris pH 7.5) and frozen at -20°C and capture ELISAs were performed as follows.

For Capture ELISAs, 250 ng of an ammonium sulfate IgG cut of goat polyclonal antibody to gp120SF2/env2-3 was used to coat each well of a 96-well plate (Corning, Corning, NY). Serial dilutions of gp120/SF2 protein (MID 167) were used to set the quantitation curve from which expression of US4 or SF162 gp120 proteins from transfection supernatant and lysates were calculated. Samples were screened undiluted and, optionally, by serial 2-fold dilutions. A human polyclonal antibody to HIV-1 gp120/SF2 was used to detect bound gp120 envelope protein, followed by horse-radish peroxidase (HRP)-labeled goat anti-human IgG conjugates. TMB (Pierce, Rockford, IL) was used as the substrate and the reaction is terminated by addition of 4N H<sub>2</sub>SO<sub>4</sub>. The reaction was quantified by measuring the optical density (OD) at 450 nm. The intensity of the color is directly

proportional to the amount of HIV gp120 antigen in a sample. Purified SF2 gp120 protein was diluted and used as a standard.

- 5 The results of the transient expression assays are presented in Tables 3 and 4. Table 3 depicts transient expression in 293 cells transfected with a pCMVKm2 vector carrying the Env cassette of interest. Table 4 depicts transient expression in RD cells transfected with a pCMVKm2 vector carrying the Env cassette of interest.

5

Table 3

Native (N) Synthetic (S)	Cell Line	Total sup (ng)	Sup fold increase (S v. N)	Total cell lysate (ng)	Cell lysate fold increase (S v. N)	Total (ng)	Total fold increase (S v. N)
N-gp120.US4	RD	87		<1		88	
S-gp120.modUS4	RD	690	8	2	5	693	8
N-gp140.US4	RD	526		0		526	
S-gp140.modUS4	RD	1305	2	1	2	1306	2
S-gp140mut.modUS4	RD	35	N/A	25	N/A	60	N/A
S-gp140TM.modUS4	RD	0	N/A	5	N/A	5	N/A
N-gp160.US4	RD	0		8		8	
S-gp160.modUS4	RD	0	0	30	4	30	4

Table 4

CHO Cell Lines Expression Level of US4 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
gp120.modUS4	1	3.2 $\mu$ M	250-450
	2	1.6 $\mu$ M	350-450
	3	200nM	230-580
	4	200nM	300-500
gp140.modUS4	1	1 $\mu$ M	155-300
	2	1 $\mu$ M	100-260
	3	1 $\mu$ M	200-430
gp140.mut. modUS4	1	1 $\mu$ M	110-270
	2	1 $\mu$ M	100-235
	3	1 $\mu$ M	100-220
gp140.modUS4 .delV1/V2	1	50nM	313-587**
	2	50nM	237-667**
	3	50nM	492-527**
gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
	2	50nM	82-318**
	3	50nM	204-385**

\*All samples measured at T-75 flask stage unless otherwise indicated

\*\*at 24 well and 6 well plate stages

\*\*\*in a three liter bioreactor perfusion culture this clone yielded approximately 2-5  $\mu$ g/ml.



The data showed that the synthetic Env and expression cassettes provided a significant increase in production of their protein products, relative to the native (HIV-1SF162 or US4 wild-type) sequences, when  
5 expressed in a variety of cell lines.

C. CHO Cell line Env expression data

Chinese hamster ovary (CHO) cells were transfected with plasmid DNA encoding the synthetic HIV-1 gp120 or  
10 gp140 proteins (e.g., pESN2dhfr or pCMVIII vector backbone) using Mirus TransIT-LT1 polyamine transfection reagent (Pan Vera) according to the manufacturers instructions and incubated for 96 hours. After 96 hours, media was changed to selective media (F12 special with  
15 250 µg/ml G418) and cells were split 1:5 and incubated for an additional 48 hours. Media was changed every 5-7 days until colonies started forming at which time the colonies were picked, plated into 96 well plates and screened by gp120 Capture ELISA. Positive clones were  
20 expanded in 24 well plates and screened several times for Env protein production by Capture ELISA, as described above. After reaching confluency in 24 well plates, positive clones were expanded to T25 flasks (Corning, Corning, NY). These were screened several times after  
25 confluency and positive clones were expanded to T75 flasks.

Positive T75 clones were frozen in LN2 and the highest expressing clones amplified with 0-5 µM methotrexate (MTX) at several concentrations and plated in  
30 100mm culture dishes. Plates were screened for colony formation and all positive clones were again expanded as described above. Clones were expanded and amplified and screened at each step by gp120 capture ELISA. Positive clones were frozen at each methotrexate level. Highest

producing clones were grown in perfusion bioreactors (3L, 100L) for expansion and adaptation to low serum suspension culture conditions for scale-up to larger bioreactors.

- 5           Tables 5 and 6 show Capture ELISA data from CHO cells transfected with pCMVIII vector carrying a cassette encoding synthetic HIV-US4 and SF162 Env polypeptides (e.g., mutated cleavage sites, modified codon usage and/or deleted hypervariable regions). Thus, stably
- 10   transfected CHO cell lines which express Env polypeptides (e.g., gp120, gp140-monomeric, and gp140-oligomeric) have been produced.

Table 5

CHO Cell Lines Expression Level of US4 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level' (ng/ml)
gp120.modUS4	1	3.2 $\mu$ M	250-450
	2	1.6 $\mu$ M	350-450
	3	200nM	230-580***
	4	200nM	300-500
gp140.modUS4	1	1 $\mu$ M	155-300
	2	1 $\mu$ M	100-260
	3	1 $\mu$ M	200-430
gp140.mut. modUS4	1	1 $\mu$ M	110-270
	2	1 $\mu$ M	100-235
	3	1 $\mu$ M	100-220
gp140.modUS4 .delV1/V2	1	50nM	313-587**
	2	50nM	237-667**
	3	50nM	492-527**
gp140.mut. modUS4.delV1 /V2	1	50nM	46-328**
	2	50nM	82-318**
	3	50nM	204-385**

\*All samples measured at T-75 flask stage unless otherwise indicated

\*\*at 24 well and 6 well plate stages

\*\*\*in a three liter bioreactor perfusion culture this clone yielded approximately 2-5  $\mu$ g/ml.

Table 6

CHO Cell Lines Expression Level of SF162 Envelope Constructs			
Constructs	CHO Clone #	MTX Level	Expression Level* (ng/ml)
gp120.modSF162	1	0	755-2705
	2	0	928-1538
	3	0	538-1609
gp140.modSF162	1	20 nM	180-350
gp140.mut. modSF162	1	20 nM	164-451
	2	20 nM	188-487
	3	20 nM	233-804
gp120.modSF162 .delV2	1	800nM	528-1560
	2	800nM	487-1878
	3	800nM	589-1212
gp140.modSF162 .delV2	1	800nM	300-600
	2	800nM	200-400
	3	800nM	200-500
gp140.mut. modSF162.delV2	1	800nM	300-700
	2	400nM	1161
	3	800nM	400-600
	4	400nM	1600-2176

\*All samples measured at T-75 flask stage unless otherwise indicated

The results presented above demonstrate the ability of the constructs of the present invention to provide expression of Env polypeptides in CHO cells. Production of polypeptides using CHO cells provides (i) correct glycosylation patterns and protein conformation (as determined by binding to panel of MAbs); (ii) correct binding to CD4 receptor molecules; (iii) absence of non-

mammalian cell contaminants (e.g., insect viruses and/or cells); and (iv) ease of purification.

#### D. Tat Coding Sequences

5       The HIV-SF162 ("SF162") wild-type Tat (SEQ ID NO:85) sequences were cloned into expression vectors having the same features as the vectors into which the synthetic Tat sequences were cloned (SEQ ID NOs:87, 88 and 89).

10       Expression efficiencies for various vectors carrying the SF162 wild-type and synthetic Tat sequences are evaluated essentially as described above for Gag and Env using capture ELISAs with the appropriate anti-tat antibodies and/or CHO cell assays. Expression of the polypeptides encoded by the synthetic cassettes is  
15       improved relative to wild type.

### Example 3

#### Western Blot Analysis of Expression

##### A. Gag and Gag-Protease Coding Sequences

20       Human 293 cells were transfected as described in Example 2 with pCMV6a-based vectors containing native or synthetic Gag expression cassettes. Cells were cultivated for 60 hours post-transfection. Supernatants were prepared as described. Cell lysates were prepared  
25       as follows. The cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO) in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego,  
30       CA) were loaded with 20  $\mu$ l of supernatant or 12.5  $\mu$ l of cell lysate. A protein standard was also loaded (5  $\mu$ l, broad size range standard; BioRad Laboratories, Hercules, CA). Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad

Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer (Millipore), where the transfer was performed at 100 volts for 90 minutes. The membranes were exposed to HIV-1-positive human patient serum and immunostained using o-phenylenediamine dihydrochloride (OPD; Sigma).

The results of the immunoblotting analysis showed that cells containing the synthetic Gag expression cassette produced the expected p55 protein at higher per-cell concentrations than cells containing the native expression cassette. The Gag p55 protein was seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Gag expression cassette of the present invention. Experiments performed in support of the present invention suggest that cells containing the synthetic Gag-prot expression cassette produced the expected Gag-prot protein at comparably higher per-cell concentrations than cells containing the native expression cassette.

In addition, supernatants from the transfected 293 cells were fractionated on sucrose gradients. Aliquots of the supernatant were transferred to Polyclear™ ultra-centrifuge tubes (Beckman Instruments, Columbia, MD), under-laid with a solution of 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 28,000 rpm in a Beckman SW28 rotor. The resulting pellet was suspended in PBS and layered onto a 20-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor.

The gradient was then fractionated into approximately 10 x 1 ml aliquots (starting at the top, 20%-end, of the gradient). Samples were taken from

fractions 1-9 and were electrophoresed on 8-16% SDS polyacrylamide gels. Fraction number 4 (the peak fraction) corresponds to the expected density of Gag protein VLPs. The supernatants from 293/synthetic Gag cells gave much stronger p55 bands than supernatants from 293/native Gag cells, and, as expected, the highest concentration of p55 in either supernatant was found in fraction 4.

These results demonstrate that the synthetic Gag expression cassette provides superior production of both p55 protein and VLPs, relative to the native Gag coding sequences.

#### B. Env Coding Sequences

Human 293 cells were transfected as described in Example 2 with pCMVKm2-based; pCMVlink-based; p-CMVII-based or pESN2-based vectors containing native or synthetic Env expression cassettes. Cells were cultivated for 48 or 60 hours post-transfection. Cell lysates and supernatants were prepared as described (Example 2). Briefly, the cells were washed once with phosphate-buffered saline, lysed with detergent [1% NP40 (Sigma Chemical Co., St. Louis, MO)] in 0.1 M Tris-HCl, pH 7.5], and the lysate transferred into fresh tubes. SDS-polyacrylamide gels (pre-cast 8-16%; Novex, San Diego, CA) were loaded with 20  $\mu$ l of supernatant or 12.5  $\mu$ l of cell lysate. A protein molecular weight standard and an HIV SF2 gp120 positive control protein (5  $\mu$ l, broad size range standard; BioRad Laboratories, Hercules, CA) were also loaded. Electrophoresis was carried out and the proteins were transferred using a BioRad Transfer Chamber (BioRad Laboratories, Hercules, CA) to Immobilon P membranes (Millipore Corp., Bedford, MA) using the transfer buffer recommended by the manufacturer

(Millipore), where the transfer was performed at 100 volts for 90 minutes. The membranes were then reacted against polyclonal goat anti-gp120SF2/env2-3 anti-sera, followed by incubation with swine anti-goat IgG-  
5 peroxidase (POD) (Sigma, St. Louis, MO). Bands indicative of binding were visualized by adding DAB with hydrogen peroxide which deposits a brown precipitate on the membranes.

The results of the immunoblotting analysis showed  
10 that cells containing the synthetic Env expression cassette produced the expected Env gp proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at higher per-cell concentrations than cells containing the native  
15 expression cassette. The Env proteins were seen in both cell lysates and supernatants. The levels of production were significantly higher in cell supernatants for cells transfected with the synthetic Env expression cassette of the present invention.

#### 20 C. Tat Coding Sequences

Human 293 cells are transfected as described in Example 2 with various vectors containing native or synthetic Tat expression cassettes. Cells are cultivated  
25 and isolated proteins analyzed as described above. Immunoblotting analysis shows that cells containing the synthetic Tat expression cassette produced the expected Tat proteins of the predicted molecular weights as determined by mobilities in SDS-polyacrylamide gels at  
30 higher per-cell concentrations than cells containing the native expression cassette.



Example 4Purification of Env polypeptidesA. Purification of Oligomeric gp140

Purification of oligomeric gp140 (o-gp140 US4) was  
5 conducted essentially as shown in Figure 60. For the  
experiments described herein, o-gp140 refers to  
oligomeric gp140 in either native or modified (e.g.,  
optimized expression sequences, deleted, mutated,  
truncated, etc.) form. Briefly, concentrated (30-50X)  
10 supernatants obtained from CHO cell cultures were loaded  
onto an anion exchange (DEAE) column which removed DNA  
and other serum proteins. The eluted material was loaded  
onto a ceramic hydroxyapatite column (CHAP) which bound  
serum proteins but not HIV Env proteins. The flow-  
15 through from the DEAE and CHAP columns was loaded onto a  
Protein A column as a precautionary step to remove any  
remaining serum immunoglobulins. The Env proteins in the  
flow-through were then captured using the lectin  
gluconanthus navalis (GNA, Vector Labs, Burlingame, CA).  
20 GNA has high affinity for mannose rich carbohydrates such  
as Env. The Env proteins were then eluted with GNA  
substrate. To remove other highly glycosylated proteins,  
a cation exchange column (SP) was used to purify  
gp140/gp120. In a final step, which separates gp120 from  
25 o-gp140, a gel filtration column was used to separate  
oligomers from monomers. Sizing and chromatography  
analysis of the final product revealed that this strategy  
lead to the successful isolation of oligomeric gp140.

B. Purification of gp120

30 Purification of gp120 was conducted essentially as  
previously described for other Env proteins. Briefly,  
concentrated supernatants obtained from CHO cell cultures  
were loaded onto an anion exchange (DEAE) column which

removed DNA and other serum proteins. The eluted material was loaded onto a ceramic hydroxyapatite column (CHAP) which bound serum proteins but not HIV Env proteins. The flow-through from the CHAP column was  
5 loaded a cation exchange column (SP) where the flow-through was discarded and the bound fraction eluted with salt. The eluted fraction(s) were loaded onto a Suprose 12/Superdex 200 Tandem column (Pharmacia-Upjohn, Uppsala, Sweden) from which purified gp120 was obtained. Sizing  
10 and chromatography analysis of the final product revealed that this strategy successfully purified gp120 proteins.

#### Example 5

##### Analysis of Purified Env Polypeptides

###### 15 A. Analysis of o-gp140

It is well documented that HIV Env protein binds to CD4 only in its correct conformation. Accordingly, the ability of o-gp140 US4 polypeptides, produced and purified as described above, to bind CD4 cells was  
20 tested. O-gp140 US4 was incubated for 15 minutes with FITC-labeled CD4 at room temperature and loaded onto a Biosil 250 (BioRad) size exclusion column using Waters HPLC. CD4-FITC has the longest retention time (2.67 minutes), followed by CD4-FITC-gp120 (2.167 min). The  
25 shortest retention time (1.9 min) was observed for CD4-FITC-o-gp140 US4 indicating that, as expected, o-gp140 US4 binds to CD4 forming a large complex which reduces retention time on the column. Thus, the o-gp140 US4 produced and purified as described above is of the  
30 correct size and conformation.

In addition, the US4 o-gp140, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible

site, the V3 loop and oligomer-specific gp41 epitope. O-gp140 bound strongly to these antibodies, indicating that the purified protein retains its structural integrity.

5    B. Analysis of gp120

As described above, CD4-FITC binds gp120, as demonstrated by the decreased retention time on the HPLC column. Thus, US4 gp120 purified by the above method retains its conformational integrity. In addition, the properties of purified gp120 can be tested by examining its integrity and identity on western blots, as well as, by examining protein concentration, pH, conductivity, endotoxin levels, bioburden and the like. US4 gp120, purified as described above, was also tested for its ability to bind to a variety of monoclonal antibodies with known epitope specificities for the CD4 binding site, the CD4 inducible site, the V3 loop and oligomer-specific gp41 epitope. The pattern of mAb binding to gp120 indicated that the purified protein retained its structural integrity, for example, the purified gp120 did not bind the mAb having the oligomer-specific gp41 epitope (as expected).

Example 6

25    Electron Microscopic Evaluation of VLP Production

The cells for electron microscopy were plated at a density of 50-70% confluence, one day before transfection. The cells were transfected with 10 µg of DNA using transfection reagent LT1 (Panvera) and incubated for 5 hours in serum-reduced medium (see Example 2). The medium was then replaced with normal medium (see Example 2) and the cells were incubated for 14 hours (COS-7) or 40 hours (CHO). After incubation the cells were washed twice with PBS and fixed with 2%

glutaraldehyde. Electron microscopy was performed by Prof. T.S. Benedict Yen, Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. The magnification was 100,000X.

Figures 3A and 3B show micrographs of CHO cells transfected with pCMVKM2 carrying the synthetic Gag expression cassette (SEQ ID NO:5) or carrying the Gag-prot expression cassette (SEQ ID NO:79). In the figure, free and budding immature virus-like-particles (VLP) of the expected size (100 nm) are seen for the Gag expression cassette (Figure 3A) and both immature and mature VLPs are seen for the Gag-prot expression cassette (Figure 3B). COS-7 cells transfected with the same vector have the same expression pattern. VLP can also be found intracellularly in CHO and COS-7 cells.

Native and synthetic Gag expression cassettes were compared for their associated levels of VLP production when used to transfect human 293 cells. The comparison was performed by density gradient ultracentrifugation of cell supernatants and Western-blot analysis of the gradient fractions. There was a clear improvement in production of VLPs when using the synthetic Gag construct.

#### Example 7

#### Expression of Virus-like Particles in the Baculovirus System

##### A. Expression of Native HIV p55 Gag

To construct the native HIV p55 Gag baculovirus shuttle vector, the prototype SF2 HIV p55 plasmid, pTM1-

Gag (Selby M.J., et al., *J Virol.* 71(10):7827-7831, 1997), was digested with restriction endonucleases *Nco*I and *Bam*HI to extract a 1.5 Kb fragment that was subsequently subcloned into pAcC4 (*Bio/Technology* 6:47-55, 1988), a derivative of pAc436. Generation of the recombinant baculovirus was achieved by co-transfecting 2 µg of the HIV p55 Gag pAcC4 shuttle vector with 0.5 µg of linearized, *Autographa californica* baculovirus (AcNPV) wild-type viral DNA into *Spodoptera frugiperda* (Sf9) cells (Kitts, P.A., Ayres M.D., and Possee R.D., *Nucleic Acids Res.* 18:5667-5672, 1990). The isolation of recombinant virus expressing HIV p55 Gag was performed according to standard techniques (O'Reilly, D.R., L.K. Miller, and V. A. Luckow, *Baculovirus Expression Vector: A Laboratory Manual*, W.H. Freeman and Company, New York, 1992).

Expression of the HIV p55 Gag was achieved using a 500 ml suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano, *Bio/Technology* 6:1506-1510, 1988) that had been infected with the HIV p55 Gag recombinant baculovirus at a multiplicity of infection (MOI) of 10. Forty-eight hours post-infection, the supernatant was separated by centrifugation and filtered through a 0.2 µm filter. Aliquots of the supernatant were then transferred to Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes, underlaid with 20% (wt/wt) sucrose, and subjected to 2 hours centrifugation at 24,000 rpm using a Beckman SW28 rotor.

The resulting pellet was suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylenediaminetetraacetic acid [EDTA]), layered onto a 20-60% (wt/wt) sucrose gradient, and subjected to 2 hours centrifugation at 40,000 rpm using a Beckman SW41ti

rotor. The gradient was then fractionated starting at the top (20% sucrose) of the gradient into approximately twelve 0.75 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining (Figure 4). Additional aliquots were subjected to refractive index analysis.

The results shown in Figure 4 indicated that the p55 Gag virus-like particles banded at a sucrose density of range of 1.15 - 1.19 g/ml with the peak at approximately 1.17 g/ml. The peak fractions were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of Tris buffer (described above). The total protein yield as estimated by Bicimchrominic Acid (BCA) (Pierce Chemical, Rockford, IL) was 1.6 mg.

#### B. Expression of Synthetic HIV p55 Gag

A baculovirus shuttle vector containing the synthetic p55 Gag sequence was constructed as follows. The synthetic HIV p55 expression cassette (Example 1) was digested with restriction enzyme *SalI* followed by incubation with T4-DNA polymerase. The resulting fragment was isolated (PCR Clean-Up™, Promega, Madison, WI) and then digested with *BamHI* endonuclease. The shuttle vector pAcC13 (Munemitsu S., et al., *Mol Cell Biol.* 10(11):5977-5982, 1990) was linearized by digestion with *EcoI*, followed by incubation with T4-DNA polymerase, and then isolated (PCR Clean-Up™). The linearized vector was digested with *BamHI*, treated with alkaline phosphatase, and isolated by size fragmentation in an agarose gel. The isolated 1.5 kb fragment was ligated with the prepared pAcC13 vector. The resulting clone was designated pAcC13-Modif.p55Gag.

The expression conditions for the synthetic HIV p55 VLPs differed from those of the native p55 Gag as follows: a culture volume of 1 liter used instead of 500 ml; *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nermerow, G.R., *BioTechnology Progress*, 9:25-30, 1993) insect cells were used instead of Sf9 insect cells; and, an MOI of 3 was instead of an MOI of 10. Experiments performed in support of the present invention showed that there was no appreciable difference in expression level between the Sf9 and Tn5 insect cells with the native p55 clone. In terms of MOI, experience with the native p55 clone suggested that an MOI of 10 resulted in higher expression (approximately 2-fold) of VLPs than a lower MOI.

The sucrose pelleting and banding methods used for the synthetic p55 VLPs were similar to those employed for the native p55 VLPs (described above), with the following exceptions: pelleted VLPs were suspended in 4 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer; and four, 20-60% sucrose gradients were used instead of a single gradient. Also, due to the high concentration of banded VLPs, further concentration by pelleting was not required. The peak fractions from all 4 gradients were simply dialyzed against PBS. The approximate density of the banded VLPs ranged from 1.23-1.28 g/ml. A total protein yield as estimated by BCA was 46 mg. Results from the sucrose gradient banding of the synthetic p55 are shown in Figure 5.

A comparison of the total amount of purified HIV p55 Gag from several preparations obtained from the two baculovirus expression cassettes has been summarized in Figure 6. The average yield from the native p55 was 3.16 mg/liter of culture (n=5, standard deviation (sd)  $\pm 1.07$ , range = 1.8-4.8 mg/L) whereas the average yield from the

synthetic p55 was more than ten-fold higher at 44.5 mg/liter of culture (n=2, sd=±6.4).

In addition to a higher total protein yield, the final product from the synthetic p55-expressed Gag consistently contained lower amounts of contaminating baculovirus proteins than the final product from the native p55-expressed Gag. This difference can be seen in the two commassie-stained gels Figures 4 and 5.

10 C. Expression of Native and Synthetic Gag-Core

Expression of the HIV p55 Gag/HCV Core 173 (SEQ ID NO:8) was achieved using a 2.5 liter suspension culture of Sf9 cells grown in serum-free medium (Miaorella, B., D. Inlow, A. Shauger, and D. Harano. 1988 Bio/Technology 6:1506-1510). The cells were infected with an HIV p55 Gag/HCV Core 173 recombinant baculovirus. Forty-eight hours post-infection, the supernatant was separated from the cells by centrifugation and filtered through a 0.2 µm filter. Aliquots of the supernatant were then transferred to a Polyclear™ (Beckman Instruments, Palo Alto, CA) ultracentrifuge tubes containing 30% (wt/wt) sucrose, and subjected to 2 hours of centrifugation at 24,000 rpm in a Beckman SW28 rotor and ultracentrifuge.

The resulting pellet was suspended in Tris buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl) and layered onto a 30-60% (wt/wt) sucrose gradient and subjected to 2 hours centrifugation at 40,000 rpm in a Beckman SW41ti rotor and ultracentrifuge. The gradient was then fractionated starting at the top (30%) of the gradient into approximately 11 x 1.0 ml aliquots. A sample of each fraction was electrophoresed on 8-16% SDS polyacrylamide gels and the resulting bands were visualized after commassie staining.



A subset of aliquots were also subjected to Western blot analysis using monoclonal antibody 76C.5EG (Steimer, K.S., et al., Virology 150:283-290, 1986) which is specific for HIV p24 (a subunit of HIV p55). The peak fractions from the sucrose gradient were pooled and concentrated by a second 20% sucrose pelleting. The resulting pellet was suspended in 1 ml of buffer Tris buffer and the total protein yield as estimated by BCA (Pierce Chemical, Rockford, IL) was ~ 1.0 mg.

The results from the SDS PAGE are shown in Figure 8 and the anti- p24 Western blot results are shown in Figure 9. Taken together, these results indicate that the HIV p55 Gag/HCV Core 173 chimeric VLPs banded at a sucrose density similar to that of the HIV p55 Gag VLPs and the visible protein band that migrated at a molecular weight of ~ 72,000 kd was reactive with the HIV p24-specific monoclonal antibody. An additional immunoreactive band at approximately 55,000 kd also appeared to be reactive with the anti-p24 antibody and may be a degradation product.

Although aliquots from the above preparation were not tested for reactivity with an HCV Core-specific antibody (an anti-CD22 rabbit serum), results from a similar preparation are shown in Figure 10 and indicate that the main HCV Core-specific reactivity migrates at an approximate molecular weight of 72,000 kd which is in accordance with the predicted molecular weight of the chimeric protein.

The expression conditions for the synthetic HIV p55 Gag/HCV Core 173 (SEQ ID NO:8) VLPs differed from those of the native p55 Gag and are as follows: a culture volume of 1 liter used instead of 2.5 liters, *Trichoplusia ni* (Tn5) (Wickham, T.J., and Nemerow, G.R. 1993 BioTechnology Progress, 9:25-30) insect cells were

used instead of Sf9 insect cells and an MOI of 3 was instead of an MOI of 10. The sucrose pelleting and banding methods used for the synthetic HIV p55 Gag/HCV Core 173 VLPs were similar to those employed for the native HIV p55 Gag/HCV Core 173 VLPs. However, differences included: pelleted VLPs were suspended in 1 ml of phosphate buffered saline (PBS) instead of 1.0 ml of the Tris buffer, and a single 20-60% sucrose gradients was used. A comparison of the total amount of purified HIV p55 Gag/HCV Core 173 from multiple preparations obtained from the two baculovirus expression cassettes showed that there was an increase in expression using the synthetic HIV p55 Gag/HCV Core 173 cassette.

D. Alternative method for the enrichment of HIV p55 Gag VLPs

In addition to purification from the media, p55 (Gag protein) expressed in baculovirus (e.g., using a synthetic expression cassette of the present invention) can also be purified as virus-like particles from the infected insect cells. For example, forty-eight hours post infection, the media and cell pellet are separated by centrifugation and the cell pellet is stored at -70°C until future use. At the time of processing, the cell pellet is suspended in 5 volumes of hypotonic lysis buffer (20 mM Tris-HCl, pH 8.2, 1 mM EGTA; 1 mM MgCl<sub>2</sub>, and Complete Protease Inhibitor® (Boehringer Mannheim Corp., Indianapolis, IN)). If needed, the cells are then dounced 8-10 times to complete cell lysis.

The lysate is then centrifuged at approximately 1000-1500 x g for 20 minutes. The supernatant is

decanted into UltraClear™ tubes, underlayered with 20% sucrose (w/w) and centrifuged at 24,000 rpm in SW28 buckets for 2 hours. The resulting pellet is suspended in Tris buffer (20 mM Tris HCl, pH 7.5, 250 mM NaCl, and 2.5 mM ethylenediamine-tetraacetic acid (EDTA) with 0.1% IGEPAL detergent (Sigma Chemical, St. Louis, MO) and 250 units/ml of benzonase (American International Chemical, Inc., Natick, MA) and incubated at 4°C for at least 30 minutes. The suspension is subsequently layered onto a 20-60% sucrose gradient and spun at 40,000 rpm using an SW41ti rotor for 20-24 hours.

After ultracentrifugation, the sucrose gradient is fractionated and aliquots run on SDS PAGE to identify peak fractions. The peak fractions are dialyzed against PBS and measured for protein content. Negatively stained electron micrographs typically show non-enveloped VLPs somewhat smaller in diameter (80-120 nm) than the budded VLPs. HIV Gag VLPs prepared in this manner are also capable of generating Gag-specific CTL responses in mice.

#### Example 8

##### In Vivo Immunogenicity of Synthetic Gag Expression

##### Cassettes

##### A. Immunization

To evaluate the possibly improved immunogenicity of the synthetic Gag expression cassettes, a mouse study was performed. The plasmid DNA, pCMVKM2 carrying the synthetic Gag expression cassette, was diluted to the following final concentrations in a total injection volume of 100 µl: 20 µg, 2 µg, 0.2 µg, and 0.02 µg. To

overcome possible negative dilution effects of the diluted DNA, the total DNA concentration in each sample was brought up to 20  $\mu$ g using the vector (pCMVCM2) alone. As a control, plasmid DNA of the native Gag expression cassette was handled in the same manner. Twelve groups of four Balb/c mice (Charles River, Boston, MA) were intramuscularly immunized (50  $\mu$ l per leg, intramuscular injection into the *tibialis anterior*) according to the schedule in Table 7.

Table 7

Group	Gag Expression Cassette	Concentration of Gag plasmid DNA ( $\mu$ g)	Immunized at time (weeks):
1	Synthetic	20	0 <sup>1</sup> , 4
2	Synthetic	2	0, 4
3	Synthetic	0.2	0, 4
4	Synthetic	0.02	0, 4
5	Synthetic	20	0
6	Synthetic	2	0
7	Synthetic	0.2	0
8	Synthetic	0.02	0
9	Native	20	0
10	Native	2	0
11	Native	0.2	0
12	Native	0.02	0

1 = initial immunization at "week 0"

Groups 1-4 were bled at week 0 (before immunization), week 4, week 6, week 8, and week 12. Groups 5-12 were bled at week 0 (before immunization) and at week 4.

### B. Humoral Immune Response

The humoral immune response was checked with an anti-HIV Gag antibody ELISAs (enzyme-linked immunosorbent assays) of the mice sera 0 and 4 weeks post immunization (groups 5-12) and, in addition, 6 and 8 weeks post immunization, respectively, 2 and 4 weeks post second immunization (groups 1-4).

The antibody titers of the sera were determined by anti-Gag antibody ELISA. Briefly, sera from immunized mice were screened for antibodies directed against the HIV p55 Gag protein. ELISA microtiter plates were coated with 0.2  $\mu$ g of HIV-1<sub>SP2</sub> p24-Gag protein per well overnight and washed four times; subsequently, blocking was done with PBS-0.2% Tween (Sigma) for 2 hours. After removal of the blocking solution, 100  $\mu$ l of diluted mouse serum was added. Sera were tested at 1/25 dilutions and by serial 3-fold dilutions, thereafter. Microtiter plates were washed four times and incubated with a secondary, peroxidase-coupled anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA plates were washed and 100  $\mu$ l of 3, 3', 5, 5'-tetramethyl benzidine (TMB; Pierce) was added per well. The optical density of each well was measured after 15 minutes. The titers reported are the reciprocal of the dilution of serum that gave a half-maximum optical density (O.D.). The ELISA results are presented in Table 8.

Table 8

Group	Inoculum ( $\mu$ g)	Expression cassette	Sera - Week 4 <sup>3</sup>	Sera - Week 6	Sera - Week 8
1	20	S <sup>1</sup> - gag	98	455	551
2	2	S - gag	59	1408	227
3	0.	S - gag	29	186	61
4	0.02	S - gag	< 20	< 20	< 20
5	20	S - gag	67	n.a. <sup>4</sup>	n.a.
6	2	S - gag	63	n.a.	n.a.
7	0.	S - gag	57	n.a.	n.a.
8	0.02	S - gag	< 20	n.a.	n.a.
9	20	N <sup>2</sup> - gag	43	n.a.	n.a.
10	2	N - gag	< 20	n.a.	n.a.
11	0.	N - gag	< 20	n.a.	n.a.
12	0.02	N - gag	< 20	n.a.	n.a.

1 = synthetic gag expression cassette (SEQ ID NO: 4)

2 = native gag expression cassette (SEQ ID NO: 1)

3 = geometric mean antibody titer

4 = not applicable

The results of the mouse immunizations with plasmid-DNAs show that the synthetic expression cassettes provide a clear improvement of immunogenicity relative to the native expression cassettes. Also, the second boost immunization induced a secondary immune response after two weeks (groups 1-3).

### C. Cellular Immune Response

The frequency of specific cytotoxic T-lymphocytes (CTL) was evaluated by a standard chromium release assay of peptide pulsed Balb/c mouse CD4 cells. Gag expressing vaccinia virus infected CD-8 cells were used as a positive control (vvGag). Briefly, spleen cells (Effector cells, E) were obtained from the BALB/c mice immunized as described above (Table 8) were cultured, restimulated, and assayed for CTL activity against Gag

peptide-pulsed target cells as described (Doe, B., and Walker, C.M., *AIDS* 10(7):793-794, 1996). The HIV-1<sub>sf2</sub> Gag peptide used was p7g SEQ ID NO:10. Cytotoxic activity was measured in a standard <sup>51</sup>Cr release assay. Target (T)

5 cells were cultured with effector (E) cells at various E:T ratios for 4 hours and the average cpm from duplicate wells was used to calculate percent specific <sup>51</sup>Cr release. The results are presented in Table 9.

Cytotoxic T-cell (CTL) activity was measured in  
10 splenocytes recovered from the mice immunized with HIV Gag DNA (compare Effector column, Table 9, to immunization schedule, Table 8). Effector cells from the Gag DNA-immunized animals exhibited specific lysis of Gag p7g peptide-pulsed SV-BALB (MHC matched) targets cells  
15 indicative of a CTL response. Target cells that were peptide-pulsed and derived from an MHC-unmatched mouse strain (MC57) were not lysed (Table 9; MC/p7g).

Table 9

Table 9. Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 *gag* DNA

Immunization	E:T	Percent specific lysis of target cells*		
		SVBALB none	SVBALB p7g	RMA p7g
20 µg DNA gagmod	100:1	2	49	<1
	30:1	3	30	<1
	10:1	<1	14	<1
2 µg DNA gagmod	100:1	2	37	<1
	30:1	2	21	<1
	10:1	<1	13	<1
0.2 µg DNA gagmod	100:1	2	32	<1
	30:1	3	25	<1
	10:1	1	14	<1
0.02 µg DNA gagmod	100:1	1	17	<1
	30:1	1	16	<1
	10:1	1	8	<1
20 µg DNA gag native	100:1	2	49	<1
	30:1	2	24	<1
	10:1	1	12	<1
2 µg DNA gag native	100:1	<1	18	<1
	30:1	1	14	<1
	10:1	1	7	<1
0.2 µg DNA gag native	100:1	3	30	<1
	30:1	3	17	<1
	10:1	2	7	<1
0.02 µg DNA gag native	100:1	4	2	<1
	30:1	1	2	<1
	10:1	1	2	<1

\*representative results of two animals per DNA-dose; positive CTL responses are indicated by boxed data

The results of the CTL assays show increased potency of synthetic Gag expression cassettes for induction of cytotoxic T-lymphocyte (CTL) responses by DNA immunization.



Example 9In vivo Immunization with Env polypeptidesA. Immunogenicity Study of US4 o-gp140 in Ras-3c Adjuvant System

5       Studies have been conducted using rabbits immunized with US4 o-gp140 purified as described above. Studies are also underway in animals to determine immunogenicity of US4 gp120, SF162 o-gp140 and SF162 gp120.

10       Two rabbits (#1 and #2) were immunized intramuscularly at 0, 4, 12 and 24 weeks with 50  $\mu$ g of US4 o-gp140 in the Ribi™ adjuvant system (RAS-3c), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL, Ribi Immunochem, Hamilton, MT).

15       In each experiment described herein, o-gp140 can be native, mutated and/or modified. Antibody responses directed against the US4 o-gp140 protein were measured by ELISA. Results are shown in Table 10.

Table 10

Rabbit/sample	Approximate o-gp140 ELISA titer
pre-immunization	0
#1: post1 (0 week immuniz)	400
#1: post2 (4 week immuniz)	15,000
#1: post3 (12 week immuniz)	50,000
#1: post4 (24 week immuniz)	100,000
#2: post1 (0 week immuniz)	600
#2: post2 (4 week immuniz)	12,000
#2: post3 (12 week immuniz)	25,000
#2: post4 (24 week immuniz)	55,000

The avidities of antibodies directed against the US4 o-gp140 protein were measured in a similar ELISA format employing successive washes with increasing concentrations of ammonium isothiocyanate. Results are shown in Table 11.

Table 11

Time of sample	Approx. Antibody avidity (NH <sub>4</sub> HCN Conc. in M)
pre-immunization	0.02
post1 (0 week immuniz)	1.8
post2 (4 week immuniz)	3.5
post3 (12 week immuniz)	5.5
post4 (24 week immuniz)	5.1

These results show that US4 o-gp140 is highly immunogenic and able to induce substantial antibody responses after only one or two immunizations.

#### 5 B. Immunogenicity of US4 o-gp140 in MF59-based Adjuvants

Groups of 4 rabbits were immunized intramuscularly at 0, 4, 12 and 24 weeks with various doses of US4 o-gp140 protein in three different MF59-based adjuvants (MF59 is described in International Publication No. WO 90/14837 and typically contains 5% Squalene, 0.5% Tween 80, and 0.5% Span 85). Antibody titers were measured post-third by ELISA using SF2 gp120 to coat the plates. QHC is a quill-based adjuvant (Iscotek, Uppsala, Sweden). Results are shown in Table 12.

Table 12

Antigen dose ( $\mu$ g)	Adjuvant	Anti-gp120 <sub>SF2</sub> Ab GMT*
12.5	MF59	7231
25	MF59	8896
50	MF59	12822
12.5	MF59/MPL	24146
25	MF59/MPL	27199
50	MF59/MPL	23059
50	MF59/MPL/QHC	31759

\*GMT = geometric mean titer

Thus, adjuvanted o-gp140 generated antigen-specific antibodies. Further, the antibodies were shown to increased in avidity over time.

#### 30 C. Neutralizing Antibodies

Neutralizing antibodies post-third immunization were measured against HIV-1 SF2 in a T-cell line adapted virus

(TCLA) assay and against PBMC-grown HIV-1 variants SF2, SF162 and 119 using the CCR5+ CEMx174 LTR-GFP reporter cell line, 5.25 (provided by N. Landau, Salk Institute, San Diego, CA) as target cells. Results are shown in Table 13.

5

Table 13

Neutralizing antibody responses in rabbits immunized  
with o-gp140.modUS4 protein

10

15

20

25

30

Group	Animal	SF2 TCLA*	SF2 PBMC*	SF162 PBMC*	119 PBMC*
<b>Experiment 1</b>					
o-gp140/ Ras-3c 50 mg	217	>640	100%	49	17
	218	>640	96	37	29
<b>Experiment 2</b>					
o-gp140/ MF59 50 mg	792	45	71	39	26
	793	50	87	26	4
	794	59	87	13	0
	795	128	92	15	0
o-gp140/ MF59 + MPL 50 mg	804	173	91	47	18
	805	134	93	28	4
	806	N.D.**	95	49	13
	807	441	100	31	15
o-gp140/MF59 + MPL + QHC 50 mg	808	465	98	46	40
	809	496	100	44	39
	810	>640	101	27	4
	811	92	92	24	37

\*TCLA neutralizing antibody titers (50% inhibition).

\*\*Not Determined

\* % Inhibition at 1:10 dilution of sera with any detectable non-specific inhibition in pre-bleeds subtracted.

35

The above studies in rabbits indicate that the US4 o-gp140 protein is highly immunogenic. When administered with adjuvant, this protein was able to induce substantial antibody responses after only one or two immunizations.

5 Moreover, the adjuvanted o-gp140 protein was able to generate antigen-specific antibodies which increased in avidity after successive immunizations, and substantial neutralizing activity against T-cell line adapted HIV-1. Neutralizing activity was also observed against PBMC-grown  
10 primary HIV strains, including the difficult to neutralize CCR5 co-receptor (R5)-utilizing isolates, SF162 and 119.

#### Example 10

##### In Vivo Immunogenicity of Synthetic Env Expression

##### Cassettes

##### A. General Immunization Methods

To evaluate the immunogenicity of the synthetic Env expression cassettes, studies using guinea pigs, rabbits, mice, rhesus macaques and baboons were performed. The  
20 studies were structured as follows: DNA immunization alone (single or multiple); DNA immunization followed by protein immunization (boost); DNA immunization followed by Sindbis particle immunization; immunization by Sindbis particles alone.

##### B. Humoral Immune Response

The humoral immune response was checked in serum specimens from immunized animals with an anti-HIV Env antibody ELISAs (enzyme-linked immunosorbent assays) at various times post-immunization. The antibody titers of  
30 the sera were determined by anti-Env antibody ELISA as described above. Briefly, sera from immunized animals were

screened for antibodies directed against the HIV gp120 or gp140 Env protein. Wells of ELISA microtiter plates were coated

overnight with the selected Env protein and washed four  
5 times; subsequently, blocking was done with PBS-0.2% Tween  
(Sigma) for 2 hours. After removal of the blocking  
solution, 100  $\mu$ l of diluted mouse serum was added. Sera  
were tested at 1/25 dilutions and by serial 3-fold  
10 dilutions, thereafter. Microtiter plates were washed four  
times and incubated with a secondary, peroxidase-coupled  
anti-mouse IgG antibody (Pierce, Rockford, IL). ELISA  
plates were washed and 100  $\mu$ l of 3, 3', 5, 5'-tetramethyl  
benzidine (TMB; Pierce) was added per well. The optical  
density of each well was measured after 15 minutes. Titers  
15 are typically reported as the reciprocal of the dilution of  
serum that gave a half-maximum optical density (O.D.).

#### Example 11

##### DNA-immunization of Baboons Using Synthetic Gag 20 Expression Cassettes

##### A. Baboons

Four baboons were immunized 3 times (weeks 0, 4 and 8)  
bilaterally, intramuscular into the quadriceps using 1mg  
pCMVKM2.GagMod.SF2 plasmid-DNA (Example 1). The animals  
25 were bled two weeks after each immunization and a p24  
antibody ELISA was performed with isolated plasma. The  
ELISA was performed essentially as described in Example 5  
except the second antibody-conjugate was an anti-human IgG,  
g-chain specific, peroxidase conjugate (Sigma Chemical Co.,  
30 St. Louis, MD 63178) used at a dilution of 1:500. Fifty  
 $\mu$ g/ml yeast extract was added to the dilutions of plasma

samples and antibody conjugate to reduce non-specific background due to

preexisting yeast antibodies in the baboons. The antibody titer results are presented in Table 14.

5

Table 14

Immunization no.	Weeks	Antigen	wpi */ Baboon No.	Ab-titer <sup>b</sup>
10	0	gagmod DNA	0 w/219	< 10
			0 w/220	< 10
			0 w/221	< 10
			0 w/222	< 10
15	6		2 wp 1st/219	< 10
			2 wp 1st/220	< 10
			2 wp 1st/221	< 10
			2 wp 1st/222	15
20	14	gagmod DNA	2 wp 4th/219	< 10
			2 wp 4th/220	88
			2 wp 4th/221	< 10
			2 wp 4th/222	56
25	30	gagmod DNA	2 wp 5th/219	< 10
			2 wp 5th/220	391
			2 wp 5th/221	237
			2 wp 5th/222	222
30	46	gag VLP protein	2 wp 6th/219	753
			2 wp 6th/219	4330
			2 wp 6th/219	5000
			2 wp 6th/219	2881

<sup>a</sup> wpi = weeks post immunization

<sup>b</sup> geometric mean antibody titer

30

In Table 14, pre-bleed data are given as Immunization No. 0; data for bleeds taken 2 weeks post-first immunization are given as Immunization No. 1; data for bleeds taken 2 weeks post-second immunization are given as Immunization No. 2; and, data for bleeds taken 2 weeks post-third immunization are given as Immunization No. 3.

Further, lymphoproliferative responses to p24 antigen were also observed in baboons 221 and 222 two weeks post-fourth immunization (at week 14), and enhanced substantially post-boosting with VLP (at week 44 and 76).  
5 Such proliferation results are indicative of induction of T-helper cell functions.

#### B. Rhesus Macaques

The improved potency of the codon-modified *gag* expression plasmid observed in mouse and baboon studies  
10 was confirmed in rhesus macaques. Four of four macaques had detectable Gag-specific CTL after two or three 1 mg doses of modified *gag* plasmid. In contrast, in a previous study, only one of four macaques given 1 mg  
15 doses of plasmid-DNA encoding the wild-type HIV-1<sub>sf2</sub> Gag showed strong CTL activity that was not apparent until after the seventh immunization. Further evidence of the potency of the modified *gag* plasmid was the observation that CTL from two of the four rhesus macaques reacted  
20 with three nonoverlapping Gag peptide pools, suggesting that as many as three different Gag peptides are recognized and indicating that the CTL response is polyclonal. Additional quantification and specificity studies are in progress to further characterize the T  
25 cell responses to Gag in the plasmid-immunized rhesus macaques. DNA immunization of macaques with the modified *gag* plasmid did not result in significant antibody responses, with only two of four animals seroconverting at low titers. In contrast, in the same study the  
30 majority of macaques in groups immunized with p55Gag protein seroconverted and had strong Gag-specific antibody titers. These data suggest that a prime-boost



strategy (DNA-prime and protein-boost) could be very promising for the induction of a strong CTL and antibody response.

In sum, these results demonstrate that the synthetic  
5 Gag plasmid DNA is immunogenic in non-human primates. When similar experiments were carried out using wild-type Gag plasmid DNA no such induction of anti-p24 antibodies was observed after four immunizations.

10

### Example 12

#### . DNA- and Protein Immunizations of Animals Using Env Expression Cassettes and Polypeptides

##### A. Guinea Pigs

Groups comprising six guinea pigs each were  
15 immunized intramuscularly at 0, 4, and 12 weeks with plasmid DNAs encoding the gp120.modUS4, gp140.modUS4, gp140.modUS4.delV1, gp140.modUS4.delV2, gp140.modUS4.delV1/V2, or gp160.modUS4 coding sequences of the US4-derived Env. The animals were subsequently  
20 boosted at 18 weeks with a single intramuscular dose of US4 o-gp140.mut.modUS4 protein in MF59 adjuvant. Anti-gp120 SF2 antibody titers (geometric mean titers) were measured at two weeks following the third DNA  
immunization and at two weeks after the protein boost.  
25 Results are shown in Table 15.

Table 15

Group	GMT post-DNA immuniz.	GMT post-protein boost
gp120.modUS4	2098	9489
gp140.modUS4	190	5340
gp140.modUS4.delV1	341	7808
gp140.modUS4.delV2	386	8165
gp140.modUS4.delV1/V 2	664	8270
gp160.modUS4	235	9928

These results demonstrate the usefulness of the synthetic constructs to generate immune responses, as well as, the advantage of providing a protein boost to enhance the immune response following DNA immunization.

#### B. Rabbits

Rabbits were immunized intramuscularly and intradermally using a Bioject needless syringe with plasmid DNAs encoding the following synthetic SF162 Env polypeptides: gp120.modSF162, gp120.modSF162.delV2, gp140.modSF162, gp140.modSF162.delV2, gp140.mut.modSF162, gp140.mut.modSF162.delV2, gp160.modSF162, and gp160.modSF162.delV2. Approximately 1 mg of plasmid DNA (pCMVlink) carrying the synthetic Env expression cassette was used to immunize the rabbits. Rabbits were immunized with plasmid DNA at 0, 4, and 12 weeks. At two weeks after the third immunization all of the constructs were shown to have generated significant antibody titers in the test animals. Further, rabbits immunized with constructs containing deletions of the V2 region

generally generated similar antibody titers relative to rabbits immunized with the companion construct still containing the V2 region.

The nucleic acid immunizations are followed by protein boosting with o-gp140.modSF162.delV2 (0.1 mg of purified protein) at 24 weeks after the initial immunization. Results are shown in Table 16.

Table 16

Group	GMT 2wks post-2nd DNA immunization	GMT 2wks post-3rd DNA immunization	GMT 2wks post-protein boost
gp120.modSF162	4573	5899	26033
gp120.modSF162.delV2	3811	3122	29606
gp140.modSF162	1478	819	12882
gp140.modSF162.delV2	1572	819	11067
gp140.mut.modSF162	1417	788	8827
gp140.mut.modSF162.delV2	1378	1207	13301
gp160.modSF162	23	81	7050
gp160.modSF162.delV2	85	459	11568

All constructs are highly immunogenic and generate substantial antigen binding antibody responses after only 2 immunizations in rabbits.

#### C. Baboons

Groups of four baboons were immunized intramuscularly with 1 mg doses of DNA encoding different forms of synthetic US4 gp140 (see the following table) at 0, 4, 8, 12, 28, and 44 weeks. The animals were also boosted twice with US4 0-gp140 protein (gp140.mut.modUS4) at 44 and 76 weeks using MF59 as adjuvant. Results are shown in Table 17.

Table 17				
Animal	Treatment	2 Wks Post 5th DNA immuniza- tion	2 Wks post 6th DNA (plus o- gp140 prot. immuniz.)	2 Wks post 7th DNA (o-gp140 protein only)
CY 215	gp140.modUS4	8.3	446	1813
CY 216		8.3	433	1236
CY 217		68	1660	2989
CY 218		101	2556	1610
Geomean:		26.2	951.4	1812.1
CY 219	gp140.modUS4 + p55gag.SF2	8.3	8.3	421
CY 220		8.3	8.3	3117
CY 221		8.3	954	871
CY 222		8.3	71	916
Geomean:		8.3	46.5	1011.5
CY 223	gp140.mut. modUS4	41.4	10497	46432
CY 224		8.3	979	470
CY 225		135	2935	3870
CY 226		47	1209	4009
Geomean:		68.3	2457.4	4289.6
CY 227	gp140TM. modUS4	8.3	56	5001
CY 228		8.3	806	1170
CY 229		8.3	48	3402
CY 230		8.3	38	6520
GMT*:		8.3	95.3	3375.3

\*GMT = geometric mean titer

The results in Table 17 demonstrate the usefulness of the synthetic constructs to generate immune responses in primates such as baboons. In addition, all animals

showed evidence of antigen-specific (Env antigen) lymphoproliferative responses.

#### D. Rhesus Macaques

5           Two rhesus macaques (designated H445 and J408) were immunized with 1 mg of DNA encoding SF162 gp140 with a deleted V2 region (SF162.gp140.delV2) by intramuscular (IM) and intradermal (ID) routes at 0, 4, 8, and 28 weeks. Approximately 100 µg of the protein encoded by  
10   the SF162. gp140mut.delV2 construct was also administered in MF59 by IM delivery at 28 weeks.

          ELISA titers are shown in Figure 61. Neutralizing antibody activity is shown Tables 18 and 19. Neutralizing antibody activity was determined against a  
15   variety of primary HIV-1 isolates in a primary lymphocyte or "PBMC-based" assay (see the following tables). Further, the phenotypic co-receptor usage for each of the primary isolates is indicated. As can be seen in the tables neutralizing antibodies were detected against  
20   every isolate tested, including the HIV-1 primary isolates (i.e., SF128A, 92US660, 92HT593, 92US657, 92US714, 91US056, and 91US054).

Table 18					
Animal	Treatment		Bleed 0	Bleed 1	Bleed 2
	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 456	25µg 120mod DNA	(None)	8.3	45	309
EO 457			8.3	254	460
EO 458			8.3	8.3	93
EO 459			8.3	43	45
EO 460			8.3	8.3	274
EO 461	25µg 120mod DNA	25µg 120mod DNA	8.3	47	1502
EO 462			8.3	80	5776
EO 463			8.3	89	3440
EO 464			8.3	8.3	3347
EO 465			8.3	69	1127
EO 466	50µg 120mod DNA	(None)	8.3	63	102
EO 467			8.3	112	662
EO 468			8.3	94	459
EO 469			8.3	58	48
EO 470			8.3	95	355
EO 471	50µg 120mod DNA	50µg 120mod DNA	8.3	110	9074
EO 472			8.3	8.3	4897
EO 473			8.3	49	4089
EO 474			8.3	59	5280
EO 475			8.3	8.3	929
EO 476	25µg 120mod DNA	Sindbis/Env	8.3		653
EO 477			8.3	87	22675
EO 478			8.3	76	3869
EO 479			8.3		1004
EO 480			8.3	71	7080

Table 19					
	Treatment		Bleed 0	Bleed 1	Bleed 2
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 481	Sindbis/Env	(None)	8.3	8.3	8.3
EO 482			8.3	8.3	8.3
EO 483			8.3	78	103
EO 484			8.3	8.3	32
EO 485			8.3	76	207
EO 486	Sindbis/Env	Sindbis/Env	8.3	8.3	458
EO 487			8.3	8.3	345
EO 488			8.3	8.3	331
EO 489			8.3	103	111
EO 490			8.3	8.3	5636

Lymphoproliferative activity (LPA) was also determined by antigenic stimulation followed by uptake of <sup>3</sup>H-thymidine in these animals and is shown in Table 20.

Experiment 1 was performed at 14 weeks post third DNA immunization and Experiment 2 was performed at 2 weeks post fourth DNA immunization using DNA and protein. For gp120ThaiE, gp120SF2 and US4 o-gp140, appropriate background values were used to calculate Stimulation Indices (S.I.; Antigenic stimulation CPM/Background CPM).

Table 20

S.I.: Calculated as Ag CPM/Background CPM				
Animal/ exp#	gp120Thai E	gp120 SF2	env2-3SF2	o- gp140US4
J408/#1	2	1	1	5
H445/#1	1	1	1	6
J408/#2	1	1	2	3
H445/#2	0	0	3	2

As can be seen by the results presented in Table 20 lymphoproliferative responses to o-gp140.US4 antigen were also in all four animals at both experimental time points. Such proliferation results are indicative of induction of T-helper cell functions.

The results presented above demonstrate that the synthetic gp140.modSF162.delV2 DNA and protein are immunogenic in non-human primates.



## Example 13

In vitro expression of recombinant Sindbis RNA and DNA containing the synthetic Gag or Env expression cassettes

5 A. Synthetic Gag expression cassettes

To evaluate the expression efficiency of the synthetic Gag expression cassette in Alphavirus vectors, the synthetic Gag expression cassette was subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors. Specifically, a cDNA vector construct for *in vitro* transcription of Sindbis virus RNA vector replicons (pRSIN-luc; Dubensky, et al., *J Virol.* 10 70:508-519, 1996) was modified to contain a *PmeI* site for plasmid linearization and a polylinker for insertion of heterologous genes. A polylinker was generated using two 15 oligonucleotides that contain the sites *XhoI*, *PmlI*, *ApaI*, *NarI*, *XbaI*, and *NotI* (XPANXNF, SEQ ID NO:17, and XPANXNR, SEQ ID NO:18).

The plasmid pRSIN-luc (Dubensky et al., *supra*) was 20 digested with *XhoI* and *NotI* to remove the luciferase gene insert, blunt-ended using Klenow and dNTPs, and purified from an agarose gel using GeneCleanII (Biol01, Vista, CA). The oligonucleotides were annealed to each other and ligated into the plasmid. The resulting construct 25 was digested with *NotI* and *SacI* to remove the minimal Sindbis 3'-end sequence and A<sub>40</sub> tract, and ligated with an approximately 0.4 kbp fragment from PKSSIN1-BV (WO 97/38087). This 0.4 kbp fragment was obtained by digestion of pKSSIN1-BV with *NotI* and *SacI*, and 30 purification after size fractionation from an agarose gel. The fragment contained the complete Sindbis virus 3'-end, an A<sub>40</sub> tract and a *PmeI* site for linearization. This new vector construct was designated SINBVE.

The synthetic HIV Gag coding sequence was obtained from the parental plasmid by digestion with *EcoRI*, blunt-ending with Klenow and dNTPs, purification with GeneCleanII, digestion with *SalI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Gag coding fragment was ligated into the SINBVE vector that had been digested with *XhoI* and *PmlI*. The resulting vector was purified using GeneCleanII and designated SINBVGag. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVGag and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line as described, for example, in U.S. Patent Numbers 5,843,723 and 5,789,245, and then administered *in vivo* as described..

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Gag gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINBVGag with *SalI* and *XhoI*, purification using GeneCleanII of the Gag-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Gag, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested with the Coulter p24 capture ELISA (Example 2).

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of p24 (in ng/ml) is presented in Table 21. In the table, SINGag#1 and 2 represent duplicate measurements, and SIN $\beta$ gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 21

Construct	Supernatant	Lysate
SIN $\beta$ gal RNA	0	0
SINGag#1 RNA	7 ng	Max (approx. 1 $\mu$ g)
SINGag#2 RNA	1 ng	700 ng

293 cells were transfected using LT-1 (Example 2) with recombinant Sindbis DNA. Synthetic pCMVKM2GagMod.SF2 was used as a positive control. Supernatants and lysates were collected 48h post transfection. The expression of p24 (in ng/ml) is presented in Table 22.

Table 22

Construct	Supernatant	Lysate
SINGag DNA	3	30
pCMVKM2.GagMod.SF2 DNA	32	42

The results presented in Tables 21 and 22 demonstrate that Gag proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Gag expression cassette (p55Gag.mod).

#### B. Synthetic Env expression cassettes

To evaluate the expression efficiency of the synthetic Env expression cassette in Alphavirus vectors,

synthetic Env expression cassettes were subcloned into both plasmid DNA-based and recombinant vector particle-based Sindbis virus vectors as described above for Gag.

The synthetic HIV Env coding sequence was obtained from the parental plasmid by digestion with *SalI* and *XbaI*, size fractionation on an agarose gel, and purification from the agarose gel using GeneCleanII. The synthetic Env coding fragment was ligated into the SINBVE vector that had been digested with *XhoI* and *XbaI*. The resulting vector was purified using GeneCleanII and designated SINBVEEnv. Vector RNA replicons may be transcribed *in vitro* (Dubensky et al., *supra*) from SINBVEEnv and used directly for transfection of cells. Alternatively, the replicons may be packaged into recombinant vector particles by co-transfection with defective helper RNAs or using an alphavirus packaging cell line and administered as described above for Gag.

The DNA-based Sindbis virus vector pDCMVSIN-beta-gal (Dubensky, et al., *J Virol.* 70:508-519, 1996) was digested with *SalI* and *XbaI*, to remove the beta-galactosidase gene insert, and purified using GeneCleanII after agarose gel size fractionation. The HIV Env gene was inserted into the the pDCMVSIN-beta-gal by digestion of SINBVEEnv with *XbaI* and *XhoI*, purification using GeneCleanII of the Env-containing fragment after agarose gel size fractionation, and ligation. The resulting construct was designated pDSIN-Env, and may be used directly for *in vivo* administration or formulated using any of the methods described herein.

BHK and 293 cells were transfected with recombinant Sindbis vector RNA and DNA, respectively. The supernatants and cell lysates were tested by capture ELISA.

BHK cells were transfected by electroporation with recombinant Sindbis RNA. The expression of Env (in ng/ml) is presented in Table 23. In the table, the Sindbis RNA containing synthetic Env expression cassettes are indicated and  $\beta$ gal represents a negative control. Supernatants and lysates were collected 24h post transfection.

Table 23

Construct	Supernatant (Neat)ng/ml	Lysate (1:10 dilution)ng/ml
$\beta$ gal RNA	0	0
gp140.modUS4	726	7147
gp140.modSF162	3529	7772
gp140.modUS4.delV1/V2	1738	6526
gp140.modUS4.delV2	960	3023
gp140.modSF162.delV2	2772	3359

293 cells were transfected using LT-1 mediated transfection (PanVera) with recombinant Sindbis DNA containing synthetic expression cassettes of the present invention and  $\beta$ gal sequences as a negative control. Supernatants and lysates were collected 48h post transfection. The expression of Env (in ng/ml) is presented in Table 24.

Table 24

Construct	Supernatant (Neat) ng/ml	Lysate (1:10 dilution) ng/ml
$\beta$ gal	0	0
gp140.modSF162.delV2	1977	801
gp140.modSF162	949	746

The results presented in Tables 23 and 24 demonstrated that Env proteins can be efficiently expressed from both DNA and RNA-based Sindbis vector systems using the synthetic Env expression cassettes of the present invention.

#### Example 14

##### A. In vivo Immunization with Gag-containing DNA and/or Sindbis particles

CB6F1 mice were immunized intramuscularly at 0 and 4 weeks with plasmid DNA and/or Sindbis vector RNA-containing particles each containing GagMod.SF2 sequences as indicated in Table 25. Animals were challenged with recombinant vaccinia expressing SF2 Gag at 3 weeks post second immunization (at week 7). Spleens were removed from the immunized and challenged animals 5 days later for a standard  $^{51}\text{C}$  release assay for CTL activity. Values shown in Table 25 indicate the results from the spleens of three mice from each group. The boxed values in Table 25 indicate that all groups of mice receiving immunizations with pCMVKm2.GagMod.SF2 DNA and/or SindbisGagMod.SF2 virus particles either alone or in combinations showed antigen-specific CTL activity.

Table 25

Cytotoxic T-lymphocyte (CTL) responses in mice immunized with HIV-1 gagmod DNA and Sindbis gagmod virus particles

Immunization	E:T	Percent specific lysis of target cells*		
		SVBALB none	SVBALB p7g	RMA p7g
pCMVKm2.GagMod.SF2 DNA <sup>a</sup>	100:1	5	20	1
at 0, 4 wks	25:1	5	20	<1
	6:1	4	8	<1
SindbisGagMod.SF2 virus particles <sup>b</sup>	100:1	10	49	<1
at 0, 4 weeks	25:1	7	20	<1
	6:1	5	12	<1
pCMVKm2.GagMod.SF2 DNA at 0 wks	100:1	9	58	<1
SindbisGagMod.SF2 virus particles at 4 wks	25:1	7	42	2
	6:1	4	13	<1
SindbisGagMod.SF2 virus particles at 4 wks	100:1	5	38	<1
	25:1	4	18	<1
pCMVKm2.GagMod.SF2 DNA at 0 wks	6:1	3	13	1

<sup>a</sup> 20 µg

<sup>b</sup> 10<sup>7</sup> particles

\* Challenge with recombinant vaccinia virus expressing HIV-1SF2 Gag at 3 weeks post second immunization (week 7). Spleens taken 5 days later. Ex vivo CTL assay performed by standard <sup>51</sup>Cr release assay. Values seen represent results from 3 pooled mouse spleens per group

#### B. In vivo Immunization with Env-containing DNA and/or Sindbis particles

Balb/C mice were immunized intramuscularly at 0 and 4 weeks(as shown in the following table) with plasmid DNA and/or Sindbis-virus RNA-containing particles each containing gp120.modUS4 sequences. Treatment regimes and antibody titers are shown in Table 26. Antibody titers were determined by ELISA using gp120 SF2 protein to coat the plates.

Table 26					
	Treatment		Bleed 0	Bleed 1 (8 wks)	Bleed 2 (10 wks)
Animal	1st Immunization	2nd Immunization	1st Imm'n	2nd Imm'n	2 Wks post 2nd
EO 456	25 $\mu$ g 120mod DNA	(None)	8.3	45	309
EO 457			8.3	254	460
EO 458			8.3	8.3	93
EO 459			8.3	43	45
EO 460			8.3	8.3	274
EO 461	25 $\mu$ g 120mod DNA	25 $\mu$ g 120mod DNA	8.3	47	1502
EO 462			8.3	80	5776
EO 463			8.3	89	3440
EO 464			8.3	8.3	3347
EO 465			8.3	69	1127
EO 466	50 $\mu$ g 120mod DNA	(None)	8.3	63	102
EO 467			8.3	112	662
EO 468			8.3	94	459
EO 469			8.3	58	48
EO 470			8.3	95	355
EO 471	50 $\mu$ g 120mod DNA	50 $\mu$ g 120mod DNA	8.3	110	9074
EO 472			8.3	8.3	4897
EO 473			8.3	49	4089
EO 474			8.3	59	5280
EO 475			8.3	8.3	929
EO 476	25 $\mu$ g 120mod DNA	Sindbis/Env	8.3		653
EO 477			8.3	87	22675
EO 478			8.3	76	3869
EO 479			8.3		1004
EO 480			8.3	71	7080
EO 481	Sindbis/Env	(None)	8.3	8.3	8.3
EO 482			8.3	8.3	8.3
EO 483			8.3	78	103
EO 484			8.3	8.3	32
EO 485			8.3	76	207
EO 486	Sindbis/Env	Sindbis/Env	8.3	8.3	458
EO 487			8.3	8.3	345
EO 488			8.3	8.3	331
EO 489			8.3	103	111
EO 490			8.3	8.3	5636

As can be seen from the data presented above, all of the mice generally demonstrated substantial immunological responses by bleed number 2. For Env, the best results were obtained using either (i) 50  $\mu$ g of gp120.modUS4 DNA for the first immunization followed by a second



immunization using 50  $\mu$ g of gp120.modUS4 DNA, or (ii) 25  $\mu$ g of gp120.modUS4 DNA for the first immunization followed by a second immunization using  $10^7$  pfus of Sindbis.

5 The results presented above demonstrate that the Env and Gag proteins of the present invention are effective to induce an immune response using Sindbis vector systems which include the synthetic Env (e.g., gp120.modUS4) or Gag expression cassettes.

10

#### Example 15

#### Co-Transfection of Env and Gag as Monocistronic and Bicistronic Constructs

DNA constructs encoding (i) wild-type US4 and SF162 Env polypeptides, (ii) synthetic US4 and SF162 Env polypeptides (gp160.modUS4, gp160.modUS4.delV1/V2, gp160.modSF162, and gp120.modSF162.delV2), and (iii) SF2gag polypeptide (i.e., the Gag coding sequences obtained from the SF2 variant or optimized sequences corresponding to the gagSF2 -- gag.modSF2) were prepared. These monocistronic constructs were co-transfected into 293T cells in a transient transfection protocol using the following combinations: gp160.modUS4; gp160.modUS4 and gag.modSF2; gp160.modUS4.delV1/V2; gp160.modUS4.delV1/V2 and gag.modSF2; gp160.modSF162 and gag.modSF2; gp120.modSF162.delV2 and gag.modSF2; and gag.modSF2 alone.

Further several bicistronic constructs were made where the coding sequences for Env and Gag were under the control of a single CMV promoter and, between the two coding sequences, an IRES (internal ribosome entry site (EMCV IRES); Kozak, M., Critical Reviews in Biochemistry and Molecular Biology 27(45):385-402, 1992; Witherell, G.W., et al., Virology 214:660-663, 1995) sequence was

introduced after the Env coding sequence and before the Gag coding sequence. Those constructs were as follows:  
gp160.modUS4.gag.modSF2, SEQ ID NO:73 (Figure 61);  
gp160.modUSF162.gag.modSF2, SEQ ID NO:74 (Figure 62);  
5 gp160.modUS4.delV1/V2.gag.modSF2, SEQ ID NO:75 (Figure 63); and gp160.modSF162.delV2.gag.modSF2, SEQ ID NO:76 (Figure 64).

Supernatants from cell culture were filtered through 0.45  $\mu$ m filters then ultracentrifuged for 2 hours at  
10 24,000 rpm (140,000Xg) in an SW28 rotor through a 20% sucrose cushion. The pelleted materials were suspended and layered on a 20-60% sucrose gradient and spun for 2 hours at 40,000 rpm (285,000Xg) in an SW41Ti rotor. Gradients were fractionated into 1.0 ml samples. A total  
15 of 9-10 fractions were typically collected from each DNA transfection group.

The fractions were tested for the presence of the Env and Gag proteins (across all fractions). These results demonstrated that the appropriate proteins were  
20 expressed in the transfected cells (i.e., if an Env coding sequence was present the corresponding Env protein was detected; if a Gag coding sequence was present the corresponding Gag protein was detected).

Virus like particles (VLPs) were known to be present  
25 through a selected range of sucrose densities. Chimeric virus like particles (VLPs) were formed using all the tested combinations of constructs containing both Env and Gag. Significantly more protein was found in the supernatant collected from the cells transfected with  
30 "gp160.modUS4.delV1/V2 and gag.modSF2" than in all the other supernatants.

Western blot analysis was also performed on sucrose gradient fractions from each transfection. The results show that bicistronic plasmids gave lower amounts of VLPs

than the amounts obtained using co-transfection with monocistronic plasmids.

In order to verify the production of chimeric VLPs by these cell lines the following electron microscopic analysis was carried out.

293T cells were plated at a density of 60-70% confluence in 100 mm dishes on the day before transfection. The cells were transfected with 10 µg of DNA in transfection reagent LT1 (Panvera Corporation, 545 Science Dr., Madison, WI). The cells were incubated overnight in reduced serum medium (opti-MEM, Gibco-BRL, Gaithersburg, MD). The medium was replaced with 10% fetal calf serum, 2% glutamine in IMDM in the morning of the next day and the cells were incubated for 65 hours. Supernatants and lysates were collected for analysis as described above (see Example 2).

The fixed, transfected 293T cells and purified ENV-GAG VLPs were analyzed by electron microscopy. The cells were fixed as follows. Cell monolayers were washed twice with PBS and fixed with 2% glutaraldehyde. For purified VLPs, gradient peak fractions were collected and concentrated by ultracentrifugation (24,000 rpm) for 2 hours. Electron microscopic analysis was performed by Prof. T.S. Benedict Yen (Veterans Affairs, Medical Center, San Francisco, CA).

Electron microscopy was carried out using a transmission electron microscope (Zeiss 10c). The cells were pre-stained with osmium and stained with uranium acetate and lead citrate. Immunostaining was performed to visualize envelope on the VLP. The magnification was 100,000X.

Figures 65A-65F show micrographs of 293T cells transfected with the following constructs: Figure 65A, gag.modSF2; Figure 65B, gp160.modUS4; Figure 65C,

gp160.modUS4.delV1/V2.gag.modSF2 (bicistronic Env and  
Gag); Figures 65D and 65E, gp160.modUS4.delV1/V2 and  
gag.modSF2; and Figure 65F, gp120.modSF162.delV2 and  
gag.modSF2. In the figures, free and budding immature  
5 virus-like-particles (VLPs) of the expected size  
(approximately 100 nm) decorated with the Env protein  
were seen. In sum, gp160 polypeptides incorporate into  
Gag VLPs when constructs were co-transfected into cells.  
The efficiency of incorporation is 2-3 fold higher when  
10 constructs encoding V-deleted Env polypeptides from high  
synthetic expression cassettes are used.

Although preferred embodiments of the subject  
invention have been described in some detail, it is  
understood that obvious variations can be made without  
15 departing from the spirit and the scope of the invention  
as defined by the appended claims.

What Is Claimed Is:

1. An expression cassette, comprising  
5 a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:20.  
10
2. The expression cassette of claim 1, comprising, a polynucleotide sequence encoding a polypeptide including an HIV Gag polypeptide, wherein the polynucleotide sequence encoding said Gag polypeptide  
15 comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:9.
3. The expression cassette of claim 1, wherein said polynucleotide sequence encoding a polypeptide including  
20 an HIV Gag polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:4.
4. The expression cassette of claim 1, wherein said  
25 polynucleotide sequence further includes a polynucleotide sequence encoding an HIV protease polypeptide.
5. The expression cassette of claim 4, wherein the  
30 nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:5, SEQ ID NO:78, and SEQ ID NO:79.
6. The expression cassette of claim 1, wherein said

polynucleotide sequence further includes a polynucleotide sequence encoding an HIV reverse transcriptase polypeptide.

5           7. The expression cassette of claim 6, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:80, SEQ ID NO:81, SEQ ID NO:82, SEQ ID NO:83, and SEQ  
10 ID NO:84.

          8. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV tat polypeptide.

15           9. The expression cassette of claim 8, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 and SEQ ID NO:89.

          10. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV polymerase  
25 polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:6.

30           11. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HIV polymerase polypeptide, wherein (i) the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90%

sequence identity to the sequence presented as SEQ ID NO:4, and (ii) wherein the sequence is modified by deletions of coding regions corresponding to reverse transcriptase and integrase.

5

12. The expression cassette of claim 11, wherein said polynucleotide sequence preserves T-helper cell and CTL epitopes.

10

13. The expression cassette of claim 1, wherein said polynucleotide sequence further includes a polynucleotide sequence encoding an HCV core polypeptide, wherein the nucleotide sequence encoding said polypeptide comprises a sequence having at least 90% sequence identity to the sequence presented as SEQ ID NO:7.

15

14. An expression cassette, comprising a polynucleotide sequence encoding a polypeptide including an HIV *Env* polypeptide, wherein the polynucleotide sequence encoding said *Env* polypeptide comprises a sequence having at least 90% sequence identity to SEQ ID NO:71 (Figure 58) or SEQ ID NO:72 (Figure 59).

20

15. The expression cassette of claim 14, wherein said *Env* polypeptide includes sequences flanking a V1 region but has a deletion in the V1 region itself.

25

16. The expression cassette of claim 15, wherein the polynucleotide sequence encoding the polypeptide comprises the sequence presented as SEQ ID NO:65 (Figure 52 gp160.modUS4.delV1).

30

17. The expression cassette of claim 14, wherein

said Env polypeptide includes sequences flanking a V2 region but has a deletion in the V2 region itself.

18. The expression cassette of claim 17, wherein  
5 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:60 (Figure 47); and SEQ ID NO:66 (Figure 53).

19. The expression cassette of claim 17, wherein  
10 the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:34 (Figure 20); SEQ ID NO:37 (Figure 24); SEQ ID NO:40 (Figure 27); SEQ ID NO:43 (Figure 30); SEQ ID NO:46 (Figure 33); SEQ ID NO:49 (Figure 36); and SEQ ID NO:76  
15 (Figure 64).

20. The expression cassette of claim 14, wherein  
said Env polypeptide includes sequences flanking a V1/V2 region but has a deletion in the V1/V2 region itself.  
20

21. The expression cassette of claim 20, wherein  
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:59 (Figure 46); SEQ ID NO:61 (Figure 48); SEQ ID NO:67  
25 (Figure 54); and SEQ ID NO:75 (Figure 63).

22. The expression cassette of claim 20, wherein  
the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:35  
30 (Figure 21); SEQ ID NO:38 (Figure 25); SEQ ID NO:41 (Figure 28); SEQ ID NO:44 (Figure 31); SEQ ID NO:47 (Figure 34) and SEQ ID NO:50 (Figure 37).



23. The expression cassette of claim 14, wherein said Env polypeptide has a mutated cleavage site that prevents the cleavage of a gp140 polypeptide into a gp120 polypeptide and a gp41 polypeptide.

5

24. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:57 (Figure 44); SEQ ID NO:61 (Figure 48); and SEQ ID NO:63 (Figure 50).

10

25. The expression cassette of claim 23, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

15

20

26. The expression cassette of claim 14, wherein said Env polypeptide includes a gp160 Env polypeptide or a polypeptide derived from a gp160 Env polypeptide.

25

27. The expression cassette of claim 26, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:64 (Figure 51); SEQ ID NO:65 (Figure 52); SEQ ID NO:66 (Figure 53); SEQ ID NO:67 (Figure 54); SEQ ID NO:68 (Figure 55); SEQ ID NO:75 (Figure 63); and SEQ ID NO:73 (Figure 61).

30

28. The expression cassette of claim 26, wherein

the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:48 (Figure 35); SEQ ID NO:49 (Figure 36); SEQ ID NO:50 (Figure 37); SEQ ID NO:76 (Figure 64); and SEQ ID NO:74 (Figure 62).

29. The expression cassette of claim 14, wherein said *Env* polypeptide includes a gp140 *Env* polypeptide or a polypeptide derived from a gp140 *Env* polypeptide.

10

30. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:56 (Figure 43); SEQ ID NO:57 (Figure 44); SEQ ID NO:58 (Figure 45); SEQ ID NO:59 (Figure 46); SEQ ID NO:60 (Figure 47); SEQ ID NO:61 (Figure 48); SEQ ID NO:62 (Figure 49); and SEQ ID NO:63 (Figure 50).

15

31. The expression cassette of claim 29, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:36 (Figure 23); SEQ ID NO:37 (Figure 24); SEQ ID NO:38 (Figure 25); SEQ ID NO:39 (Figure 26); SEQ ID NO:40 (Figure 27); SEQ ID NO:41 (Figure 28); SEQ ID NO:42 (Figure 29); SEQ ID NO:43 (Figure 30); SEQ ID NO:44 (Figure 31); SEQ ID NO:45 (Figure 32); SEQ ID NO:46 (Figure 33); and SEQ ID NO:47 (Figure 34).

25

32. The expression cassette of claim 14, wherein said *Env* polypeptide includes a gp120 *Env* polypeptide or a polypeptide derived from a gp120 *Env* polypeptide.

30

33. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:54 (Figure 41); and SEQ ID NO:55 (Figure 42).

5

34. The expression cassette of claim 32, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:33 (Figure 19); SEQ ID NO:34 (Figure 20); and SEQ ID NO:35 (Figure 21).

10

35. The expression cassette of claim 14, wherein the polynucleotide sequence encoding the polypeptide is selected from the group consisting of: SEQ ID NO:55 (Figure 42); SEQ ID NO:62 (Figure 49); SEQ ID NO:63 (Figure 50); and SEQ ID NO:68 (Figure 55).

15

36. A recombinant expression system for use in a selected host cell, comprising, an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected host cell.

20

37. The recombinant expression system of claim 36, wherein said control elements are selected from the group consisting of a transcription promoter, a transcription enhancer element, a transcription termination signal, polyadenylation sequences, sequences for optimization of initiation of translation, and translation termination sequences.

30

38. The recombinant expression system of claim 36, wherein said transcription promoter is selected from the

group consisting of CMV, CMV+intron A, SV40, RSV, HIV-Ltr, MMLV-ltr, and metallothionein.

5 39. A cell comprising an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the selected cell.

10 40. The cell of claim 39, wherein the cell is a mammalian cell.

15 41. The cell of claim 40, wherein the cell is selected from the group consisting of BHK, VERO, HT1080, 293, RD, COS-7, and CHO cells.

42. The cell of claim 41, wherein said cell is a CHO cell.

20 43. The cell of claim 39, wherein the cell is an insect cell.

44. The cell of claim 43, wherein the cell is either *Trichoplusia ni* (Tn5) or Sf9 insect cells.

25 45. The cell of claim 39, wherein the cell is a bacterial cell.

30 46. The cell of claim 39, wherein the cell is a yeast cell.

47. The cell of claim 39, wherein the cell is a plant cell.

48. The cell of claim 39, wherein the cell is an antigen presenting cell.

49. The cell of claim 48, wherein the lymphoid cell  
5 is selected from the group consisting of macrophage, monocytes, dendritic cells, B-cells, T-cells, stem cells, and progenitor cells thereof.

50. The cell of claim 39, wherein the cell is a  
10 primary cell.

51. The cell of claim 39, wherein the cell is an immortalized cell.

52. The cell of claim 39, wherein the cell is a  
15 tumor-derived cell.

53. A method for producing a polypeptide including HIV Gag polypeptide sequences, said method comprising,  
20 incubating the cells of claim 39, under conditions for producing said polypeptide.

54. A method for producing virus-like particles (VLPs), comprising,  
25 incubating the cells of claim 39, under conditions for producing said VLPs.

55. A method for producing a composition of virus-like particles (VLPs), comprising,  
30 (a) incubating the cells of claim 39, under conditions for producing said VLPs; and  
(b) substantially purifying said VLPs to produce a composition of VLPs.

56. A cell line useful for packaging lentivirus vectors, comprising

5 suitable host cells that have been transfected with an expression vector containing an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the host cell.

10 57. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of any of claims 1-13.

15 58. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of claim 1-3.

20 59. The cell line of claim 56, wherein suitable host cells have been transfected with an expression vector containing the expression cassette of claim 14-35.

60. A gene delivery vector for use in a Mammalian subject, comprising

25 a suitable gene delivery vector for use in said subject, wherein the vector comprises an expression cassette of any of claims 1-35, and wherein said polynucleotide sequence is operably linked to control elements compatible with expression in the subject.

30 61. A method of DNA immunization of a subject, comprising,

introducing a gene delivery vector of claim 60 into said subject under conditions that are compatible with expression of said expression cassette in said subject.

62. The method of claim 61, wherein said gene delivery vector is a nonviral vector.

5 63. The method of claim 61, wherein said vector is delivered using a particulate carrier.

64. The method of claim 63, wherein said vector is coated on a gold or tungsten particle and said coated  
10 particle is delivered to said subject using a gene gun.

65. The method of claim 63, wherein said vector is encapsulated in a liposome preparation.

15 66. The method of claim 61, wherein said vector is a viral vector.

67. The method of claim 66, wherein said viral vector is a retroviral vector.

20 68. The method of claim 67, wherein said viral vector is a lentiviral vector.

69. The method of claim 61, wherein said subject is  
25 a mammal.

70. The method of claim 69, wherein said mammal is a human.

30 71. A method of generating an immune response in a subject, comprising  
transfecting cells of said subject a gene delivery vector of claim 60, under conditions that permit the expression of said polynucleotide and production of said

polypeptide, thereby eliciting an immunological response to said polypeptide.

5        72. The method of claim 71, wherein said vector is a nonviral vector.

73. The method of claim 72, wherein said vector is delivered using a particulate carrier.

10       74. The method of claim 73, wherein said vector is coated on a gold or tungsten particle and said coated particle is delivered to said vertebrate cell using a gene gun.

15       75. The method of claim 73, wherein said vector is encapsulated in a liposome preparation.

20       76. The method of claim 71, wherein said vector is a viral vector.

77. The method of claim 76, wherein said viral vector is a retroviral vector.

25       78. The method of claim 77, wherein said viral vector is a lentiviral vector.

79. The method of claim 71, wherein said subject is a mammal.

30       80. The method of claim 79, wherein said mammal is a human.

81. The method of claim 71, wherein said transfecting is done ex vivo and said transfected cells



are reintroduced into said subject.

82. The method of claim 71, wherein said transfecting is done *in vivo* in said subject.

5

83. The method of claim 71, where said immune response is a humoral immune response.

84. The method of claim 71, where said immune response is a cellular immune response.

10

85. A gene delivery vector comprising an alphavirus vector construct, wherein said alphavirus construct comprises an expression cassette according to any one of claims 1 through 35.

15

86. The gene delivery vector of claim 85, wherein the alphavirus vector construct is a cDNA vector construct.

20

87. The gene delivery vector of claim 85, wherein the alphavirus comprises a recombinant alphavirus particle preparation.

88. The gene delivery vector of claim 85, wherein the vector comprises a eukaryotic layered vector initiation system.

25

89. A method of stimulating an immune response in a subject comprising administering the gene delivery vector of any one of claims 85 through 88 in an amount effective to stimulate an immune response in said subject.

30

90. The method of claim 89, wherein the gene

delivery vector is administered intramuscularly, intramucosally, intranasally, subcutaneously, intradermally, transdermally, intravaginally, intrarectally, orally or intravenously.

5

1 / 131

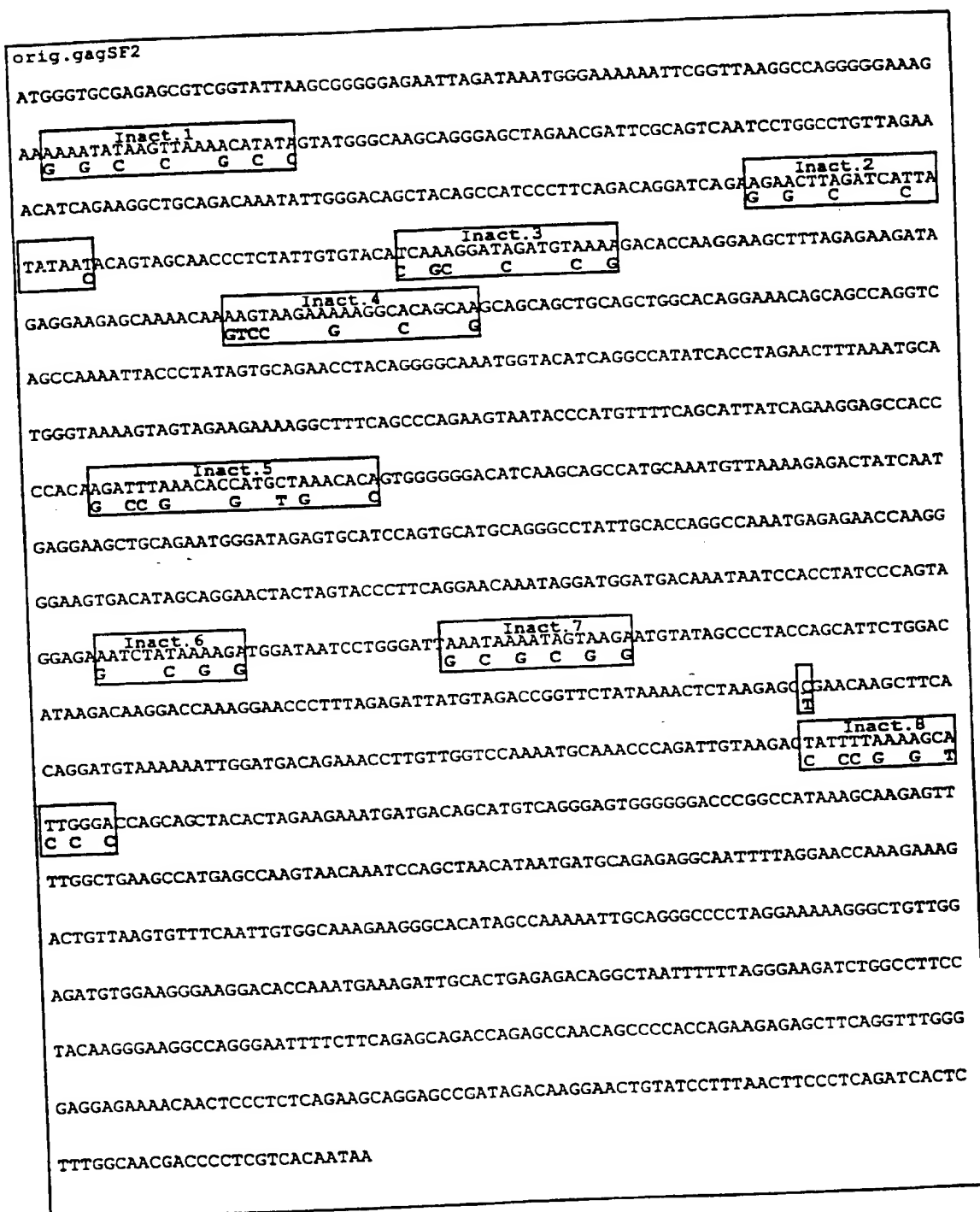


FIG. 1

SUBSTITUTE SHEET (RULE 26)

native HIV-1SF2 gag-protease 2 / 131

From here codon optimization + inactivation (GP1) and (GP2)

ATGGGTGCGAGAGCGTCGGTATTAAGCGGGGAGAAATTAGATAAATGGGAAAAAATTCGGTTAAGGCCAGGGGGAAAG

Inact. 1  
AAAAATATAAGTTAAAAATATATGATATGGGCAAGCAGGGAGCTAGAACGATTGCGAGTCAATCCTGGCCTGTTAGAA  
G G C C G C C

ACATCAGAAGGCTGCAGACAAATATTGGGACAGCTACAGCCATCCCTTCAGACAGGATCAGAGAACTTAGATCATTAA  
Inact. 2  
G G C C C

Inact. 2  
TATAATACAGTAGCAACCTCTATTGTGTACATCAAGGATAGATGTAAAGACACCAAGGAAGCTTTAGAGAAGATA  
C GC C C G

Inact. 4  
GAGGAAGAGCAAAACAAAGTAAGAAAAAGGCACAGCAAGCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC  
GTCC G C G

AGCCAAAATTACCTTATAGTGCAGAACCTACAGGGGCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA

TGGGTAAAAGTAGTAGAAGAAAAGGCTTTTTCAGCCCAGAAGTAATACCCATGTTTTTCAGCATTATCAGAAGGAGCCACC

Inact. 5  
CCACAGATTTTAAACACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAATGTTAAAAGAGACTATCAAT  
G CC G G T G C

GAGGAAGCTGCAGAAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCTATTGCACCAGGCCAAATGAGAGAACCAAGG

GGAAGTGACATAGCAGGAACCTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTA

Inact. 6  
GGAGAATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGATGTATAGCCCTACCAGCATTCTGGAG  
C G G G C G C G G

ATAAGACAAGGACCAAGGAACCTTTAGAGATTATGTAGACCGGTTCTATAAACTCTAAGAGCCCAACAAGCTTCA

Inact. 7  
CAGGATGTAAAAAATTGGATGACAGAAACCTTGTGGTCCAAAATGCAAAACCAGATTGTAAGAGCTATTTTAAAGCA  
C CC G G T

Inact. 7  
TTGGGACCGAGCAGCTACACTAGAGAAATGATGACAGCATGTCAGGGAGTGGGGGGACCCGGCCATAAAGCAAGAGTT  
C C C G

Inact. 8  
TTGGGTGAAGCCATGAGCCAAAGTAACAATCCAGCTTACATAATGATGCAGAGAGGCAATTTTAGGAACCAAGAAAG  
C G G G G G C G G C

Inact. 9  
ACTGTTAAGTGTTCATTTGTTGGCAAGAAGGGGCATAGCCAAAATTCAGGGCCCTAGGAAGAAGGGCTGTTGG  
C C C CC GG C C CC

AGATGTGAAGGGAAGGACACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGGAAGATCTGGCCTTCC

From here no changes to native sequence (GP1) and (GP2)

TACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAAAGCCACCAGAGAGCTTCAGGTTTGGG

GAGGAGAAAACAACTCCCTCTCAGAACGAGGAGCCGATAGACAAGGAAGTGTATCCTTTAACTTCCCTCAGATCACTC

Inact. 11 or only inactivation (GP1)  
TTTGCAACGACCCCTCGTCACATAAAGGATAGGGGGCAACTAAAGGAAGCTCTATTAGATACAGGAGCAGATGATA  
G C G G C G C G

Inact. 12  
CAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAACCAAAAATGATAGGGGGAATTGGAGGTTTATCAAAGTAA  
G C G C C G G

Inact. 12  
GACAGTACGATCAGATACCTGTAGAAATCTGTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCA  
G C

Inact. 13  
ACATAATTGGAAGAAATCTGTTGACTCAGATTGGTTGTACTTAAATTTCCCATAGTCTTATTGAAGTGTACCAG  
C C C C C G C C

Inact. 14  
TAAAATTAAAGCCAGGAATGGATGGCCAAAAGTAAAGCAATGGCCATTGTAA  
G G G G G C C G C

FIG. 2

SUBSTITUTE SHEET (RULE 26)



FIG. 3A

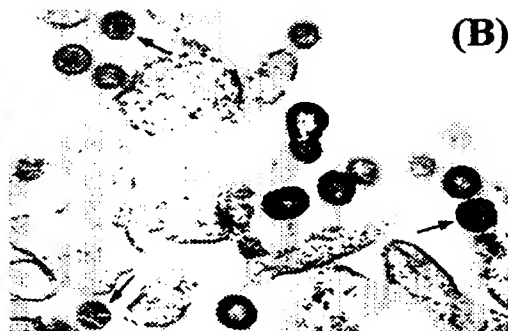


FIG. 3B

4 / 131

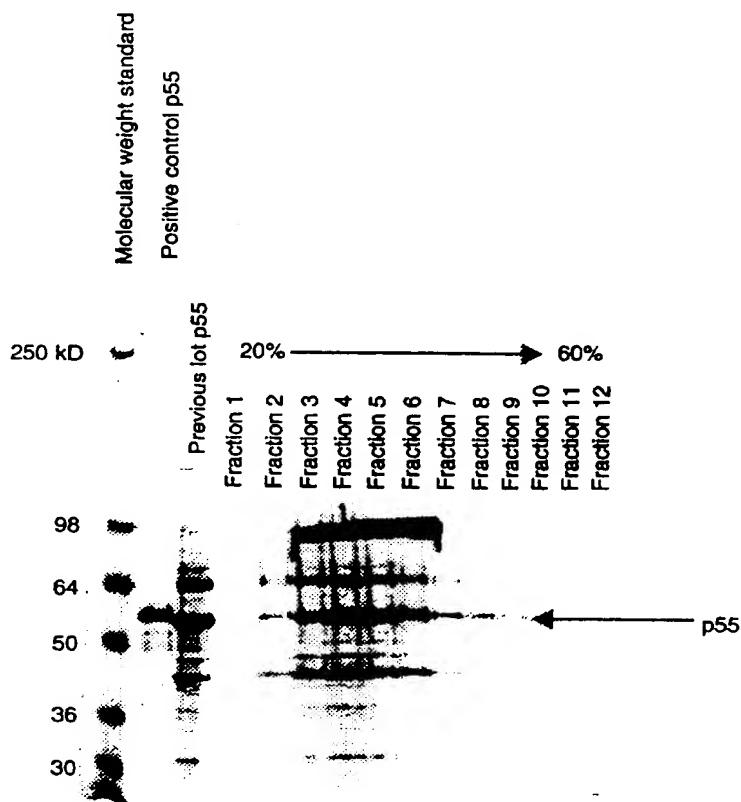


FIG. 4

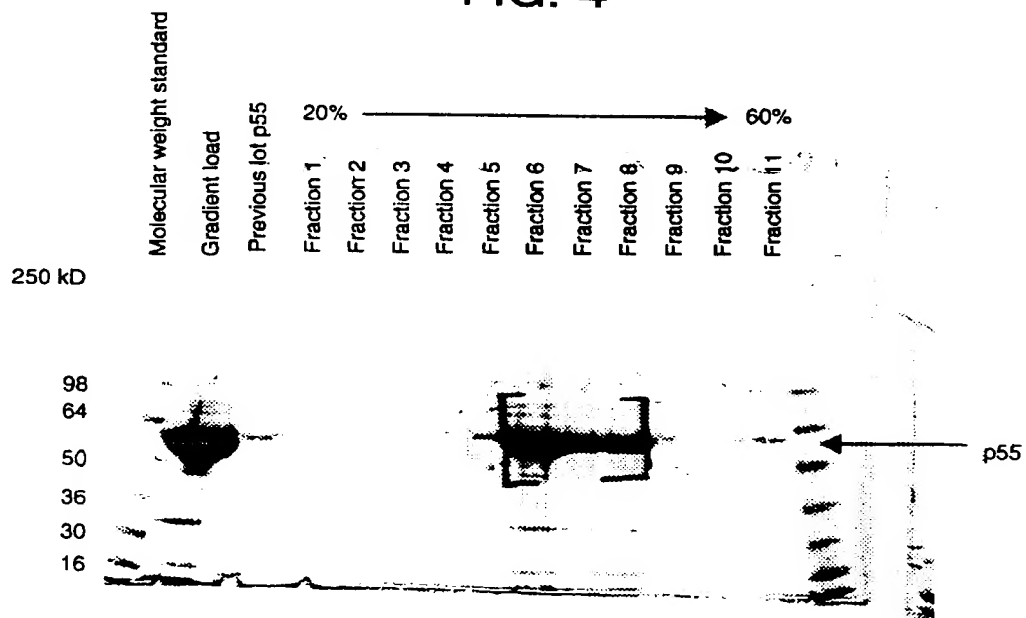


FIG. 5

SUBSTITUTE SHEET (RULE 26)

5 / 131

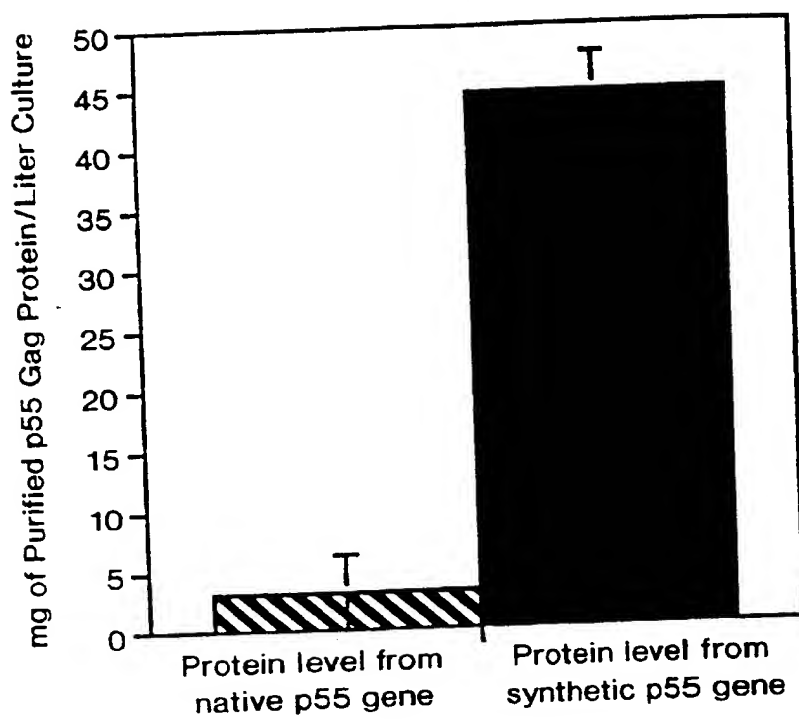


FIG. 6

SUBSTITUTE SHEET (RULE 26)

6 / 131

GagPol.ModSF	1	ATGGGCGCCC	10	GGCGGCGCT	20	GCTGAGCGG	30	GGGAGCTGG	40	ACAAAGTGGGA	50
GagProt.Mods	1	ATGGGCGCCC		GGCGGCGCT		GCTGAGCGG		GGGAGCTGG		ACAAAGTGGGA	50
Gag.ModSF2	1	ATGGGCGCCC	60	GGCGGCGCT	70	GCTGAGCGG	80	GGGAGCTGG	90	ACAAAGTGGGA	100
GagPol.ModSF	51	GAAGATCCGC	110	CTGGGCGCG	120	GGGCAAGAA	130	GAAGTACAAG	140	CTGAAGCACA	150
GagProt.Mods	51	GAAGATCCGC		CTGGGCGCG		GGGCAAGAA		GAAGTACAAG		CTGAAGCACA	150
Gag.ModSF2	51	GAAGATCCGC	110	CTGGGCGCG	120	GGGCAAGAA	130	GAAGTACAAG	140	CTGAAGCACA	150
GagPol.ModSF	101	TCGTGTGGG	160	CAGCCGCGG	170	CTGAGCGCT	180	TCGCCGTGAA	190	CCCCGGCCTG	200
GagProt.Mods	101	TCGTGTGGG		CAGCCGCGG		CTGAGCGCT		TCGCCGTGAA		CCCCGGCCTG	200
Gag.ModSF2	101	TCGTGTGGG	160	CAGCCGCGG	170	CTGAGCGCT	180	TCGCCGTGAA	190	CCCCGGCCTG	200
GagPol.ModSF	151	CTGGAGACCA	210	GGGAGGCTG	220	CCGCCAGATC	230	CTGGGCCAGC	240	TGCAGCCCCAG	250
GagProt.Mods	151	CTGGAGACCA		GGGAGGCTG		CCGCCAGATC		CTGGGCCAGC		TGCAGCCCCAG	250
Gag.ModSF2	151	CTGGAGACCA	210	GGGAGGCTG	220	CCGCCAGATC	230	CTGGGCCAGC	240	TGCAGCCCCAG	250
GagPol.ModSF	201	CCTGCAGACC	260	GGCAGGAGG	270	AGCTGGCGAG	280	CCTGTACAAC	290	ACCGTGGCCA	300
GagProt.Mods	201	CCTGCAGACC		GGCAGGAGG		AGCTGGCGAG		CCTGTACAAC		ACCGTGGCCA	300
Gag.ModSF2	201	CCTGCAGACC	260	GGCAGGAGG	270	AGCTGGCGAG	280	CCTGTACAAC	290	ACCGTGGCCA	300
GagPol.ModSF	251	CCCTGTACTG	310	CGTGACCCAG	320	CGCATCGACG	330	TCAAGGACAC	340	CAAGGAGGCC	350
GagProt.Mods	251	CCCTGTACTG		CGTGACCCAG		CGCATCGACG		TCAAGGACAC		CAAGGAGGCC	350
Gag.ModSF2	251	CCCTGTACTG	310	CGTGACCCAG	320	CGCATCGACG	330	TCAAGGACAC	340	CAAGGAGGCC	350
GagPol.ModSF	301	CTGGAGAGA	360	TCGAGGAGGA	370	GCAGAACAAAG	380	TCCAAGAAAG	390	AGGCCCCAGCA	400
GagProt.Mods	301	CTGGAGAGA		TCGAGGAGGA		GCAGAACAAAG		TCCAAGAAAG		AGGCCCCAGCA	400
Gag.ModSF2	301	CTGGAGAGA	360	TCGAGGAGGA	370	GCAGAACAAAG	380	TCCAAGAAAG	390	AGGCCCCAGCA	400
GagPol.ModSF	351	GGCCGCCGCC	410	GCCGCCGGCA	420	CCGGCAACAG	430	CAGCCAGGTG	440	AGCCAGAACT	450
GagProt.Mods	351	GGCCGCCGCC		GCCGCCGGCA		CCGGCAACAG		CAGCCAGGTG		AGCCAGAACT	450
Gag.ModSF2	351	GGCCGCCGCC	410	GCCGCCGGCA	420	CCGGCAACAG	430	CAGCCAGGTG	440	AGCCAGAACT	450
GagPol.ModSF	401	ACCCCATCGT	460	GCAGAACCTG	470	CAGGGCCAGA	480	TGTTGCACCA	490	GGCCATCAGC	500
GagProt.Mods	401	ACCCCATCGT		GCAGAACCTG		CAGGGCCAGA		TGTTGCACCA		GGCCATCAGC	500
Gag.ModSF2	401	ACCCCATCGT	460	GCAGAACCTG	470	CAGGGCCAGA	480	TGTTGCACCA	490	GGCCATCAGC	500

FIG. 7A

SUBSTITUTE SHEET (RULE 26)



7 / 131

FIG. 7B

GagPol.ModSF	451	CCCCGACCC	TGAACGCTG	GGTGAAGGTG	GTGGAGGAGA	AGGCTTCAG	500
GagProt.ModS	451	CCCCGACCC	TGAACGCTG	GGTGAAGGTG	GTGGAGGAGA	AGGCTTCAG	500
Gag.ModSF2	451	CCCCGACCC	TGAACGCTG	GGTGAAGGTG	GTGGAGGAGA	AGGCTTCAG	500
GagPol.ModSF	501	CCCCGAGGTG	ATCCCCATGT	TCAGCGCCCT	GAGCGAGGGC	GCCACCCCCC	550
GagProt.ModS	501	CCCCGAGGTG	ATCCCCATGT	TCAGCGCCCT	GAGCGAGGGC	GCCACCCCCC	550
Gag.ModSF2	501	CCCCGAGGTG	ATCCCCATGT	TCAGCGCCCT	GAGCGAGGGC	GCCACCCCCC	550
GagPol.ModSF	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GCGGCCACCA	GCCCGCCATG	600
GagProt.ModS	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GCGGCCACCA	GCCCGCCATG	600
Gag.ModSF2	551	AGGACCTGAA	CACGATGTTG	AACACCGTGG	GCGGCCACCA	GCCCGCCATG	600
GagPol.ModSF	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GCCGCCGAGT	GGGACCGCGT	650
GagProt.ModS	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GCCGCCGAGT	GGGACCGCGT	650
Gag.ModSF2	601	CAGATGCTGA	AGGAGACCAT	CAACGAGGAG	GCCGCCGAGT	GGGACCGCGT	650
GagPol.ModSF	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGGCCAGATG	GCGGAGCCCC	700
GagProt.ModS	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGGCCAGATG	GCGGAGCCCC	700
Gag.ModSF2	651	GCACCCCGTG	CACGCCGGCC	CCATCGCCCC	CGGCCAGATG	GCGGAGCCCC	700
GagPol.ModSF	701	GCGGCAGCGA	CATCGCCGGC	ACCACGAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagProt.ModS	701	GCGGCAGCGA	CATCGCCGGC	ACCACGAGCA	CCCTGCAGGA	GCAGATCGGC	750
Gag.ModSF2	701	GCGGCAGCGA	CATCGCCGGC	ACCACGAGCA	CCCTGCAGGA	GCAGATCGGC	750
GagPol.ModSF	751	TGGATGACCA	ACAAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAAGCGGTG	800
GagProt.ModS	751	TGGATGACCA	ACAAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAAGCGGTG	800
Gag.ModSF2	751	TGGATGACCA	ACAAACCCCC	CATCCCCGTG	GGCGAGATCT	ACAAAGCGGTG	800
GagPol.ModSF	801	GATCATCCTG	GGCCTGAACA	AGATCGTGCG	GATGTACAGC	CCCACGAGCA	850
GagProt.ModS	801	GATCATCCTG	GGCCTGAACA	AGATCGTGCG	GATGTACAGC	CCCACGAGCA	850
Gag.ModSF2	801	GATCATCCTG	GGCCTGAACA	AGATCGTGCG	GATGTACAGC	CCCACGAGCA	850
GagPol.ModSF	851	TCCTGGACAT	CCGCCAGGGC	CCCAGGAGC	CCTTCCCGCA	CTACGTGGAC	900
GagProt.ModS	851	TCCTGGACAT	CCGCCAGGGC	CCCAGGAGC	CCTTCCCGCA	CTACGTGGAC	900
Gag.ModSF2	851	TCCTGGACAT	CCGCCAGGGC	CCCAGGAGC	CCTTCCCGCA	CTACGTGGAC	900

SUBSTITUTE SHEET (RULE 26)

8 / 131

GagPol.ModSF	901	CGCTTCTACA	AGACCTCTGC	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
GagProt.ModS	901	CGCTTCTACA	AGACCTCTGC	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
Gag.ModSF2	901	CGCTTCTACA	AGACCTCTGC	CGCTGAGCAG	GCCAGCCAGG	ACGTGAAGAA	950
		960	970	980	990	1000	
GagPol.ModSF	951	CTGGATGACC	GAGACCTCTGC	TGGTGCAGAA	CGCCAAACCC	GA CTGCAAGA	1000
GagProt.ModS	951	CTGGATGACC	GAGACCTCTGC	TGGTGCAGAA	CGCCAAACCC	GA CTGCAAGA	1000
Gag.ModSF2	951	CTGGATGACC	GAGACCTCTGC	TGGTGCAGAA	CGCCAAACCC	GA CTGCAAGA	1000
		1010	1020	1030	1040	1050	
GagPol.ModSF	1001	CCATCCTGAA	GGCTCTCTGC	CCCGGGGCCA	CCCTGGAGGA	GATGATGACC	1050
GagProt.ModS	1001	CCATCCTGAA	GGCTCTCTGC	CCCGGGGCCA	CCCTGGAGGA	GATGATGACC	1050
Gag.ModSF2	1001	CCATCCTGAA	GGCTCTCTGC	CCCGGGGCCA	CCCTGGAGGA	GATGATGACC	1050
		1060	1070	1080	1090	1100	
GagPol.ModSF	1051	GCCTGCCAGG	GCGTGGGCGG	CCCCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
GagProt.ModS	1051	GCCTGCCAGG	GCGTGGGCGG	CCCCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
Gag.ModSF2	1051	GCCTGCCAGG	GCGTGGGCGG	CCCCGGCCAC	AAGGCCCGCG	TGCTGGCCGA	1100
		1110	1120	1130	1140	1150	
GagPol.ModSF	1101	GCGGATGAGC	CAGGTGACGA	ACCCGGCGAC	CATCATGATG	CAGCGCGGCA	1150
GagProt.ModS	1101	GCGGATGAGC	CAGGTGACGA	ACCCGGCGAC	CATCATGATG	CAGCGCGGCA	1150
Gag.ModSF2	1101	GCGGATGAGC	CAGGTGACGA	ACCCGGCGAC	CATCATGATG	CAGCGCGGCA	1150
		1160	1170	1180	1190	1200	
GagPol.ModSF	1151	ACTTCCGCAA	CCAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
GagProt.ModS	1151	ACTTCCGCAA	CCAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
Gag.ModSF2	1151	ACTTCCGCAA	CCAGCGGAAG	ACCGTCAAGT	GCTTCAACTG	CGGCAAGGAG	1200
		1210	1220	1230	1240	1250	
GagPol.ModSF	1201	GGCCACACCG	CCAGGAACTG	CCGCGCCCCC	CGCAAGAAGG	GCTGCTGGCG	1250
GagProt.ModS	1201	GGCCACACCG	CCAGGAACTG	CCGCGCCCCC	CGCAAGAAGG	GCTGCTGGCG	1250
Gag.ModSF2	1201	GGCCACACCG	CCAGGAACTG	CCGCGCCCCC	CGCAAGAAGG	GCTGCTGGCG	1250
		1260	1270	1280	1290	1300	
GagPol.ModSF	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAGA	TTGCACTGAG	AGACAGGCTA	1300
GagProt.ModS	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAGA	TTGCACTGAG	AGACAGGCTA	1300
Gag.ModSF2	1251	CTGCGGCCGC	GAAGGACACC	AAATGAAAGA	TTGCACTGAG	AGACAGGCTA	1300
		1310	1320	1330	1340	1350	
GagPol.ModSF	1301	ATTTTCTAGG	GAAGATCTGG	CCTTCCTACA	AGGGAAGGCC	AGGGAATTTT	1350
GagProt.ModS	1301	ATTTTCTAGG	GAAGATCTGG	CCTTCCTACA	AGGGAAGGCC	AGGGAATTTT	1350
Gag.ModSF2	1301	ATTTTCTAGG	GAAGATCTGG	CCTTCCTACA	AGGGAAGGCC	AGGGAATTTT	1350

FIG. 7C

SUBSTITUTE SHEET (RULE 26)

9 / 131

GagPol.ModSF	1351	CTTCAGAGCA	1360	GACCAGAGCC	1370	AACAGCCCCA	1380	CCAGAAGAGA	1390	GCTTCAGGTT	1400
GagProt.ModS	1351	CTTCAGAGCA		GACCAGAGCC		AACAGCCCCA		CCAGAAGAGA		GCTTCAGGTT	1400
Gag.ModSF2	1351	CTGCAGAGCC	1410	GCCCCGAGCC	1420	CACCGCCCCC	1430	CCCGAGGAGA	1440	GCTTCGCGTT	1450
GagPol.ModSF	1401	TGGGGAGGAG	1460	AAAACAACCTC	1470	CCTCTCAGAA	1480	GCAGGAGCCG	1490	ATAGACAAGG	1500
GagProt.ModS	1401	TGGGGAGGAG		AAAACAACCTC		CCTCTCAGAA		GCAGGAGCCG		ATAGACAAGG	1500
Gag.ModSF2	1401	CGGCGAGGAG	1460	AAGACCAACC	1470	CCAGCCAGAA	1480	GCAGGAGCC	1490	ATCGACAAGG	1500
GagPol.ModSF	1451	AACTGTATCC	1510	TTTAACTTCC	1520	CTCAGATCAC	1530	TCCTTGGCAA	1540	CGACCCCTCG	1550
GagProt.ModS	1451	AACTGTATCC		TTTAACTTCC		CTCAGATCAC		TCCTTGGCAA		CGACCCCTCG	1550
Gag.ModSF2	1451	AGCTGTATCC	1510	CCTGACCAGC	1520	CTGCGGAGCC	1530	TGTTCCGCAA	1540	CGACCCCTCG	1550
GagPol.ModSF	1501	TCACAGTAAG	1560	GATCGGCGGC	1570	CAGCTCAAGG	1580	AGGCGTGCT	1590	CGACACCGGC	1600
GagProt.ModS	1501	TCACAGTAAG		GATCGGCGGC		CAGCTCAAGG		AGGCGTGCT		CGACACCGGC	1600
Gag.ModSF2	1501	AGCCAGTAA	1560	CCGTGCTGGA	1570	GGAGATGAAC	1580	CTGCCCCGCA	1590	AGTGGAAAGCC	1600
GagPol.ModSF	1551	GCCGACGACA	1610	CCGTGCTGGA	1620	GGGGCTTCAT	1630	CAAGGTGCGG	1640	CAGTACGACC	1650
GagProt.ModS	1551	GCCGACGACA		CCGTGCTGGA		GGGGCTTCAT		CAAGGTGCGG		CAGTACGACC	1650
Gag.ModSF2	1551	.....	1610	GCGGGGATCG	1620	GGGGCTTCAT	1630	CAAGGTGCGG	1640	CAGTACGACC	1650
GagPol.ModSF	1601	CAAGATGATC	1660	GCGGGGATCG	1670	GGCCACAAGG	1680	CCATCGGCAC	1690	CGTGCTGGTG	1700
GagProt.ModS	1601	CAAGATGATC		GCGGGGATCG		GGCCACAAGG		CCATCGGCAC		CGTGCTGGTG	1700
Gag.ModSF2	1601	.....	1660	GGAGATCTGC	1670	GGCCACAAGG	1680	CCATCGGCAC	1690	CGTGCTGGTG	1700
GagPol.ModSF	1651	AGATCCCCGT	1710	GGAGATCTGC	1720	CATCGGCCGC	1730	AACCTGCTGA	1740	CCCAGATCGG	1750
GagProt.ModS	1651	AGATCCCCGT		GGAGATCTGC		CATCGGCCGC		AACCTGCTGA		CCCAGATCGG	1750
Gag.ModSF2	1651	.....	1710	CCGTGAACAT	1720	CATCGGCCGC	1730	AACCTGCTGA	1740	CCCAGATCGG	1750
GagPol.ModSF	1701	GGCCCCACCC	1760	AACTTCCCCA	1770	TCAGCCCCAT	1780	CGAGACGGTG	1790	CCCGTGAAGC	1800
GagProt.ModS	1701	GGCCCCACCC		AACTTCCCCA		TCAGCCCCAT		CGAGACGGTG		CCCGTGAAGC	1800
Gag.ModSF2	1701	.....	1760	AACTTCCCCA	1770	TCAGCCCCAT	1780	CGAGACGGTG	1790	CCCGTGAAGC	1800
GagPol.ModSF	1751	CTGCACCCCTG	1800	AACTTCCCCA	1810	TCAGCCCCAT	1820	CGAGACGGTG	1830	CCCGTGAAGC	1840
GagProt.ModS	1751	CTGCACCCCTG		AACTTCCCCA		TCAGCCCCAT		CGAGACGGTG		CCCGTGAAGC	1840
Gag.ModSF2	1751	.....	1800	AACTTCCCCA	1810	TCAGCCCCAT	1820	CGAGACGGTG	1830	CCCGTGAAGC	1840

FIG. 7D

SUBSTITUTE SHEET (RULE 26)

10/131

GagPol.ModSF	1801	1810	1820	1830	1840	1850
GagProt.ModS	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGACCGAG
Gag.ModSF2	1801	TGAAGCCGGG	GATGGACGGC	CCCAAGGTCA	AGCAGTGGCC	CCTGTAA...
		1860	1870	1880	1890	1900
GagPol.ModSF	1851	GAGAAGATCA	AGGCCCTGGT	GGAGATCTGC	ACCGAGATGG	AGAAGGAGGG
GagProt.ModS	1851	.....	.....	.....	.....	.....
Gag.ModSF2	1851	.....	.....	.....	.....	.....
		1910	1920	1930	1940	1950
GagPol.ModSF	1901	CAAGATCAGC	AAGATCGGCC	CCGAGAACC	CTACAACACC	CCCGTGTTCG
GagProt.ModS	1901	.....	.....	.....	.....	.....
Gag.ModSF2	1901	.....	.....	.....	.....	.....
		1960	1970	1980	1990	2000
GagPol.ModSF	1951	CCATCAAGAA	GAAGGACAGC	ACCAAGTGGC	GCAAGCTGGT	GGACTTCCGC
GagProt.ModS	1951	.....	.....	.....	.....	.....
Gag.ModSF2	1951	.....	.....	.....	.....	.....
		2010	2020	2030	2040	2050
GagPol.ModSF	2001	GAGCTGAACA	AGCGCACCCA	GGACTTCTGG	GAGGTGCAGC	TGGGCATCCC
GagProt.ModS	2001	.....	.....	.....	.....	.....
Gag.ModSF2	2001	.....	.....	.....	.....	.....
		2060	2070	2080	2090	2100
GagPol.ModSF	2051	CCACCCCGCC	GGCCTGAAGA	AGAAGAAGAG	CGTGACCGTG	CTGGACCTGG
GagProt.ModS	2051	.....	.....	.....	.....	.....
Gag.ModSF2	2051	.....	.....	.....	.....	.....
		2110	2120	2130	2140	2150
GagPol.ModSF	2101	GCGACGCCTA	CTTCAGCGTG	CCCCTGGACA	AGGACTTCCG	CAAGTACACC
GagProt.ModS	2101	.....	.....	.....	.....	.....
Gag.ModSF2	2101	.....	.....	.....	.....	.....
		2160	2170	2180	2190	2200
GagPol.ModSF	2151	GCCTTCACCA	TCCCCAGCAT	CAACAACGAG	ACCCCCGGCA	TCCGCTACCA
GagProt.ModS	2151	.....	.....	.....	.....	.....
Gag.ModSF2	2151	.....	.....	.....	.....	.....
		2210	2220	2230	2240	2250
GagPol.ModSF	2201	GTACAACGTG	CTGCCCCCAGG	GCTGGAAGGG	CAGCCCCGGC	ATCTTCCAGA
GagProt.ModS	2201	.....	.....	.....	.....	.....
Gag.ModSF2	2201	.....	.....	.....	.....	.....

FIG. 7E

SUBSTITUTE SHEET (RULE 26)

11 / 131

GagPol.ModSF 2251	2260	2270	2280	2290	2300
GagProt.Mods 2251	GCAGCATGAC	CAAGATCCTG	GAGCCCTTCC	GCAAGCAGAA	CCCCGACATC
Gag.ModSF2 2251	.....	.....	.....	.....	.....
GagPol.ModSF 2301	2310	2320	2330	2340	2350
GagProt.Mods 2301	GTGATCTACC	AGTACATGGA	CGACCTGTAC	GTGGGCAGCG	ACCTGGAGAT
Gag.ModSF2 2301	.....	.....	.....	.....	.....
GagPol.ModSF 2351	2360	2370	2380	2390	2400
GagProt.Mods 2351	CGGCCAGCAC	CGCACCAAGA	TCGAGGAGCT	GCGCCAGCAC	CTGCTGCGCT
Gag.ModSF2 2351	.....	.....	.....	.....	.....
GagPol.ModSF 2401	2410	2420	2430	2440	2450
GagProt.Mods 2401	GGGGCTTCAC	CACCCCGGAC	AAGAAGCACC	AGAAGGAGCC	CCCCTTCTTG
Gag.ModSF2 2401	.....	.....	.....	.....	.....
GagPol.ModSF 2451	2460	2470	2480	2490	2500
GagProt.Mods 2451	TGGATGGCT	ACGAGCTGCA	CCCCGACAAG	TGGACCGTGC	AGCCCATCAT
Gag.ModSF2 2451	.....	.....	.....	.....	.....
GagPol.ModSF 2501	2510	2520	2530	2540	2550
GagProt.Mods 2501	GCTGCCCGAG	AAGCACAGCT	GGACCGTGAA	CGACATCCAG	AAGCTGGTGG
Gag.ModSF2 2501	.....	.....	.....	.....	.....
GagPol.ModSF 2551	2560	2570	2580	2590	2600
GagProt.Mods 2551	GCAAGCTGAA	CTGGGCCAGC	CAGATCTACG	CCGGCATCAA	GGTGAAGCAG
Gag.ModSF2 2551	.....	.....	.....	.....	.....
GagPol.ModSF 2601	2610	2620	2630	2640	2650
GagProt.Mods 2601	CTGTGCAAGC	TGCTGCCCGG	CACCAAGGCC	CTGACCCGAGG	TGATCCCCCT
Gag.ModSF2 2601	.....	.....	.....	.....	.....
GagPol.ModSF 2651	2660	2670	2680	2690	2700
GagProt.Mods 2651	GACCGAGGAG	GCCGAGCTGG	AGCTGGCCGA	GAACCCGAG	ATCCTGAAGG
Gag.ModSF2 2651	.....	.....	.....	.....	.....

FIG. 7F

SUBSTITUTE SHEET (RULE 26)

12 / 131

FIG. 7G

GagPol.ModSF 2701	2710	2720	2730	2740	2750
	AGCCCGTGCA	CGAGGTGTAC	TACGACCCCA	GCAAGGACCT	GGTGGCCGAG
GagProt.Mods 2701					2750
Gag.ModSF2 2701					2750
	2760	2770	2780	2790	2800
GagPol.ModSF 2751	ATCCAGAAGC	AGGGCCAGGG	CCAGTGGACC	TACCAGATCT	ACCAGGAGCC
GagProt.Mods 2751					2800
Gag.ModSF2 2751					2800
	2810	2820	2830	2840	2850
GagPol.ModSF 2801	CTTCAAGAAC	CTGAAGACCG	GCAAATACGC	CCCATGCGC	GGCGCCACACA
GagProt.Mods 2801					2850
Gag.ModSF2 2801					2850
	2860	2870	2880	2890	2900
GagPol.ModSF 2851	CCAACGACGT	GAAGCAGCTG	ACCGAGGCCG	TGCAGAAGGT	GAGCACCCGAG
GagProt.Mods 2851					2900
Gag.ModSF2 2851					2900
	2910	2920	2930	2940	2950
GagPol.ModSF 2901	AGCATCGTGA	TCTGGGGCAA	GATCCCCAAG	TTCAAGCTGC	CCATCCAGAA
GagProt.Mods 2901					2950
Gag.ModSF2 2901					2950
	2960	2970	2980	2990	3000
GagPol.ModSF 2951	GGAGACCTGG	GAGGCCTGGT	GGATGGAGTA	CTGGCAGGCC	ACCTGGATCC
GagProt.Mods 2951					3000
Gag.ModSF2 2951					3000
	3010	3020	3030	3040	3050
GagPol.ModSF 3001	CCGAGTGGGA	GTTCTGTGAAC	ACCCCCCCC	TGGTGAAGCT	GTGGTACCAG
GagProt.Mods 3001					3050
Gag.ModSF2 3001					3050
	3060	3070	3080	3090	3100
GagPol.ModSF 3051	CTGGAGAAGG	AGCCCATCGT	GGGCGCCGAG	ACCTTCTACG	TGGACGGCGC
GagProt.Mods 3051					3100
Gag.ModSF2 3051					3100
	3110	3120	3130	3140	3150
GagPol.ModSF 3101	CGCCCAACCGC	GAGACCAAGC	TGGGCAAGGC	CGGCTACGTG	ACCGACCGCG
GagProt.Mods 3101					3150
Gag.ModSF2 3101					3150

SUBSTITUTE SHEET (RULE 26)

13 / 131

GagPol.ModSF	3151	3160	3170	3180	3190	3200
GagProt.ModS	3151	GGCGCCAGAA	GGTGTGAGC	ATCGCCGACA	CCACCAACCA	GAAGACCGAG
Gag.ModSF2	3151	.....	.....	.....	.....	.....
GagPol.ModSF	3201	3210	3220	3230	3240	3250
GagProt.ModS	3201	CTGCAGGCCA	TCCACCTGGC	CCTGCAGGAC	AGCGGCCTGG	AGGTGAACAT
Gag.ModSF2	3201	.....	.....	.....	.....	.....
GagPol.ModSF	3251	3260	3270	3280	3290	3300
GagProt.ModS	3251	CGTGACCGAC	AGCCAGTACG	CCCTGGGCAT	CATCCAGGCC	CAGCCCCGACA
Gag.ModSF2	3251	.....	.....	.....	.....	.....
GagPol.ModSF	3301	3310	3320	3330	3340	3350
GagProt.ModS	3301	AGAGCGGAGAG	CGAGCTGGTG	AGCCAGATCA	TCGAGCAGCT	GATCAAGAAG
Gag.ModSF2	3301	.....	.....	.....	.....	.....
GagPol.ModSF	3351	3360	3370	3380	3390	3400
GagProt.ModS	3351	GAGAAGGTGT	ACCTGGCCTG	GGTGCCCGCC	CACAGGGCA	TCGGCGGCAA
Gag.ModSF2	3351	.....	.....	.....	.....	.....
GagPol.ModSF	3401	3410	3420	3430	3440	3450
GagProt.ModS	3401	CGAGCAGGTG	GACAAGCTGG	TGAGCGCCGG	CATCCGCAAG	GTGCTGTTC
Gag.ModSF2	3401	.....	.....	.....	.....	.....
GagPol.ModSF	3451	3460	3470	3480	3490	3500
GagProt.ModS	3451	TGAAACGGCAT	CGACAAGGCC	CAGGAGGAGC	ACGAGAAGTA	CCACAGGCAAC
Gag.ModSF2	3451	.....	.....	.....	.....	.....
GagPol.ModSF	3501	3510	3520	3530	3540	3550
GagProt.ModS	3501	TGGCGCGCCA	TGGCCAGCGA	CTTCAACCTG	CCCCCGTGG	TGGCCCAAGGA
Gag.ModSF2	3501	.....	.....	.....	.....	.....
GagPol.ModSF	3551	3560	3570	3580	3590	3600
GagProt.ModS	3551	GATCGTGGCC	AGCTGGGACA	AGTGCCAGCT	GAAGGGCGAG	GCCATGCACG
Gag.ModSF2	3551	.....	.....	.....	.....	.....

FIG. 7H

SUBSTITUTE SHEET (RULE 26)

14 / 131

GagPol.ModSF 3601	3610	GCCAGGTGGA	CTGCAGCCCC	GGCATCTGGC	AGCTGGACTG	CACCCACCTG	3650
GagProt.Mods 3601		.....	.....	.....	.....	.....	3650
Gag.ModSF2 3601		.....	.....	.....	.....	.....	3650
GagPol.ModSF 3651	3660	GAGGGCAAGA	TCATCCTGGT	GGCCGTGCAC	GTGGCCAGCG	GCTACATCGA	3700
GagProt.Mods 3651		.....	.....	.....	.....	.....	3700
Gag.ModSF2 3651		.....	.....	.....	.....	.....	3700
GagPol.ModSF 3701	3710	GGCCGAGGTG	ATCCCCGCGG	AGACCGGCCA	GGAGACCGCC	TACTTCCTGC	3750
GagProt.Mods 3701		.....	.....	.....	.....	.....	3750
Gag.ModSF2 3701		.....	.....	.....	.....	.....	3750
GagPol.ModSF 3751	3760	TGAAGCTGGC	CGGCCGCTGG	CCCGTGAAGA	CCATCCACAC	CGACAACGGC	3800
GagProt.Mods 3751		.....	.....	.....	.....	.....	3800
Gag.ModSF2 3751		.....	.....	.....	.....	.....	3800
GagPol.ModSF 3801	3810	AGCAACTTCA	CCAGCACCAC	CGTGAAGGCC	GCCTGCTGGT	GGGCCGGCAT	3850
GagProt.Mods 3801		.....	.....	.....	.....	.....	3850
Gag.ModSF2 3801		.....	.....	.....	.....	.....	3850
GagPol.ModSF 3851	3860	CAAGCAGGAG	TTCGGCATCC	CCTACAACCC	CCAGAGCCAG	GGCGTGGTGG	3900
GagProt.Mods 3851		.....	.....	.....	.....	.....	3900
Gag.ModSF2 3851		.....	.....	.....	.....	.....	3900
GagPol.ModSF 3901	3910	AGAGCATGAA	CAACGAGCTG	AAGAAGATCA	TCGGCCAGGT	GCGCGACCAG	3950
GagProt.Mods 3901		.....	.....	.....	.....	.....	3950
Gag.ModSF2 3901		.....	.....	.....	.....	.....	3950
GagPol.ModSF 3951	3960	GCCGAGCACC	TGAAGACCGC	CGTGCAGATG	GCGGTGTTCA	TCCACAACCTT	4000
GagProt.Mods 3951		.....	.....	.....	.....	.....	4000
Gag.ModSF2 3951		.....	.....	.....	.....	.....	4000
GagPol.ModSF 4001	4010	CAAGGGCAAG	GGCGGCATCG	GCGGCTACAG	CGCCGGCCGAG	CGCATCGTGG	4050
GagProt.Mods 4001		.....	.....	.....	.....	.....	4050
Gag.ModSF2 4001		.....	.....	.....	.....	.....	4050

FIG. 71

SUBSTITUTE SHEET (RULE 26)



15 / 131

GagPol.ModSF 4051	ACATCATCGC	4060	4070	4080	4090	4100
GagProt.ModS 4051	.....			CAGACCAAGG	AGCTGCAGAA	GCAGATCACC
Gag.ModSF2 4051	.....			.....	.....	.....
	4110		4120	4130	4140	4150
GagPol.ModSF 4101	AAGATCCAGA	ACTTCCGCGT	GTACTACCGC	GACAAACAAGG	ACCCCTGTG	
GagProt.ModS 4101	.....	.....	.....	.....	.....	
Gag.ModSF2 4101	.....	.....	.....	.....	.....	
	4160		4170	4180	4190	4200
GagPol.ModSF 4151	GAAGGGCCCC	GCCAAGCTGC	TGTGGAAGGG	CGAGGGCGCC	GTGCTGATCC	
GagProt.ModS 4151	.....	.....	.....	.....	.....	
Gag.ModSF2 4151	.....	.....	.....	.....	.....	
	4210		4220	4230	4240	4250
GagPol.ModSF 4201	AGGACAAACAG	CGACATCAAG	GTGGTGCCCC	GCCGCAAGGC	CAAGATCATC	
GagProt.ModS 4201	.....	.....	.....	.....	.....	
Gag.ModSF2 4201	.....	.....	.....	.....	.....	
	4260		4270	4280	4290	4300
GagPol.ModSF 4251	CGCGACTAGG	GCAAGCAGAT	GSCCGGCGAC	GACTGCGTGG	CCAGCCGCCA	
GagProt.ModS 4251	.....	.....	.....	.....	.....	
Gag.ModSF2 4251	.....	.....	.....	.....	.....	
	4310		4320	4330	4340	4350
GagPol.ModSF 4301	GGACGAGGAC	TAG	.....	.....	.....	
GagProt.ModS 4301	.....	.....	.....	.....	.....	
Gag.ModSF2 4301	.....	.....	.....	.....	.....	

FIG. 7J

SUBSTITUTE SHEET (RULE 26)

16 / 131

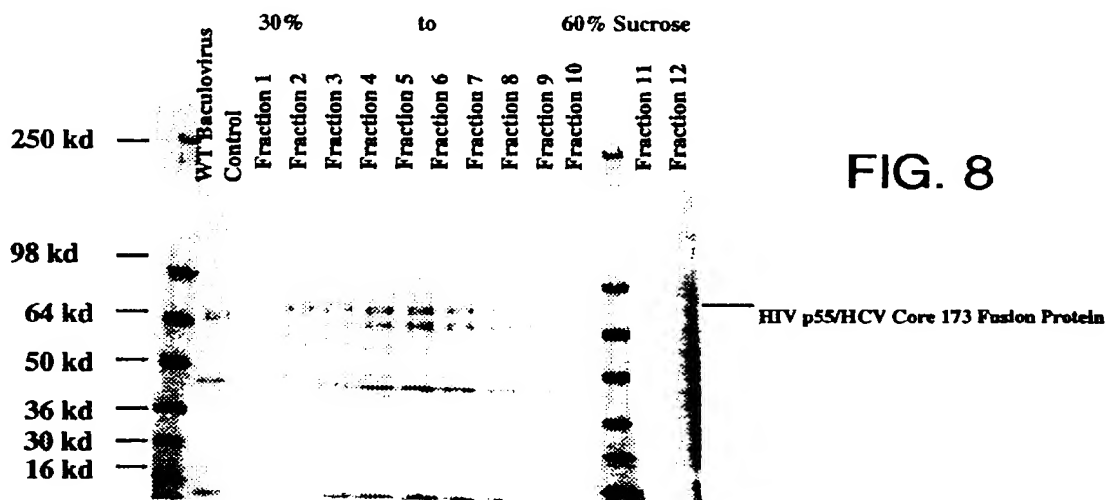


FIG. 8

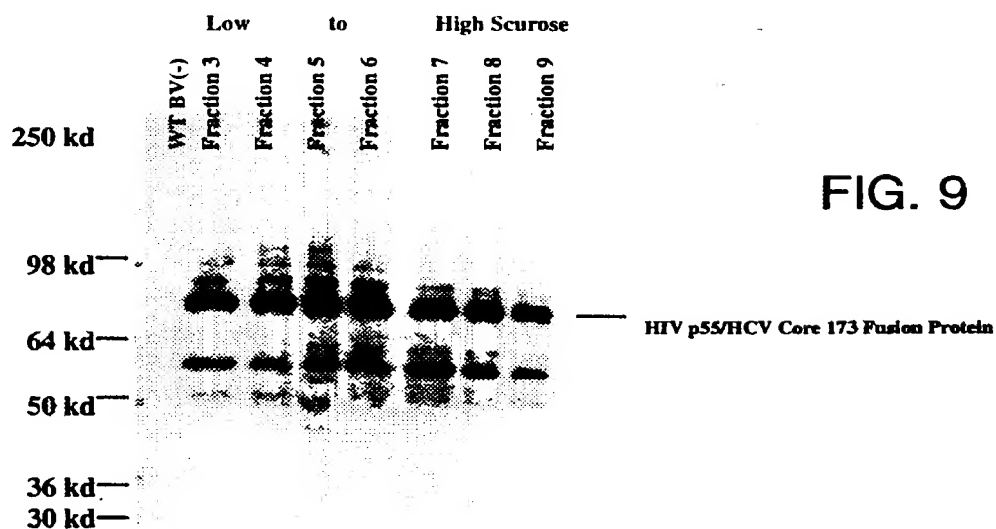


FIG. 9

SUBSTITUTE SHEET (RULE 26)

17 / 131

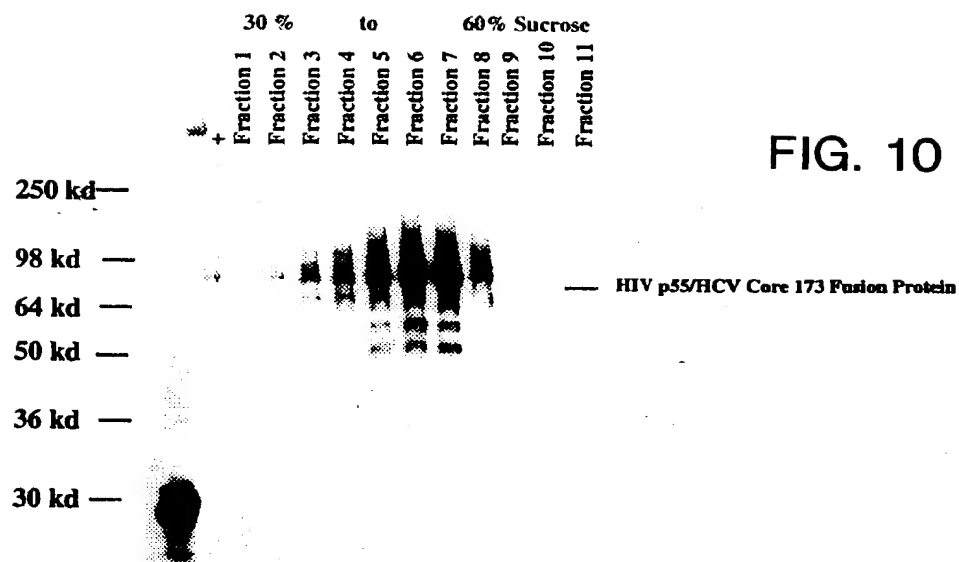


FIG. 10

SUBSTITUTE SHEET (RULE 26)

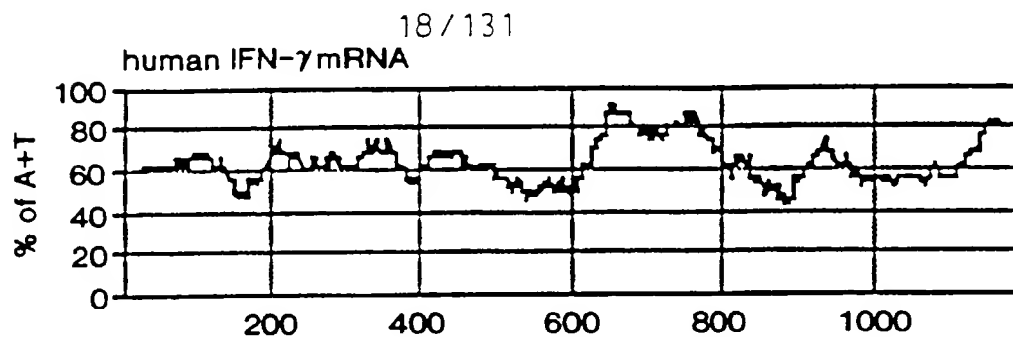


FIG. 11A

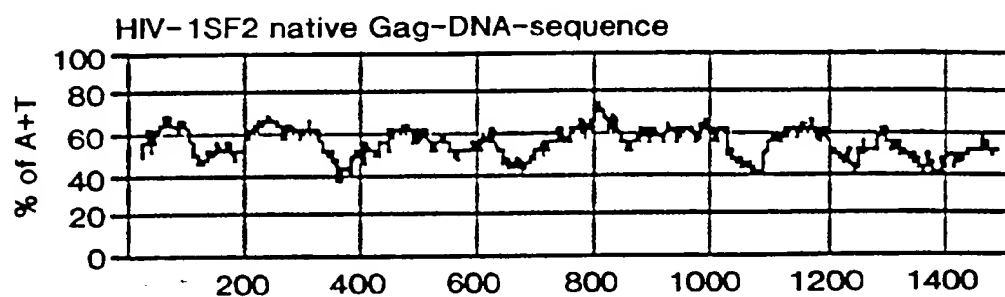


FIG. 11B

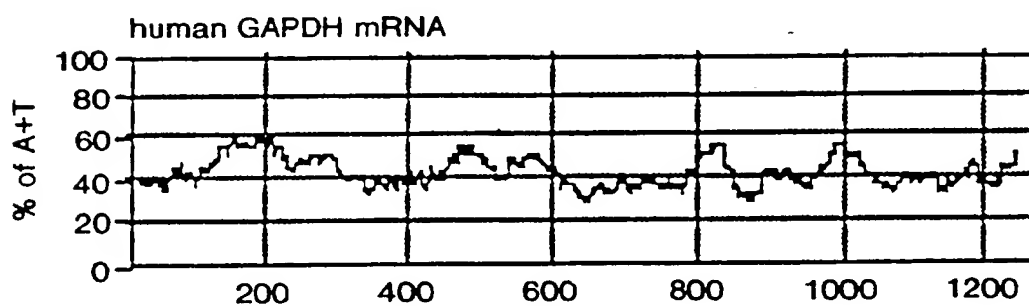


FIG. 11C

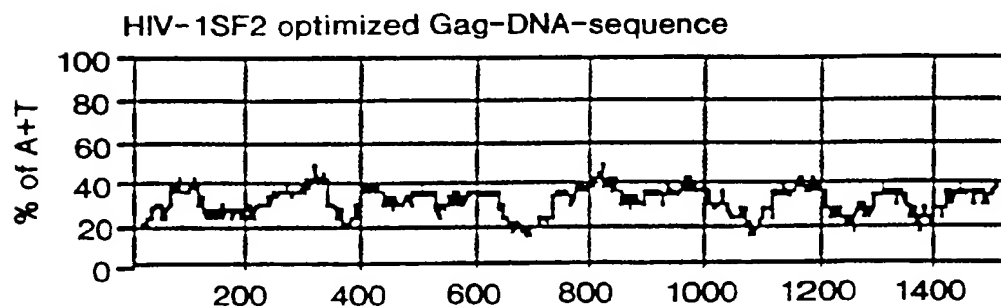


FIG. 11D

SUBSTITUTE SHEET (RULE 26)

19 / 131

native HIV-1sf2 gag-polymerase

ATGGTGCGAGAGCGCTCGGTATTAAGCGGGGAGAAATTAGATAAATGGAAAAAATTCGGTTAAGGCCAGGGGGAAG

Inact.1  
 AAAAAATATAAGTTAAAAACATATATGTTATGGCAACAGGAGCTAGAACGATTTCGAGTCAATCCTGGCCTGTTAGAA  
 G G C C G C C

Inact.2  
 ACATCAGAAGGCTGCAGACAAATATTGGGACAGCTACGCCATCCCTTCAGACAGGATCAGAAAGAACTTAGATCATT  
 G G C C

Inact.3  
 TATAATACAGTAGCAACCCCTCTATTGTGTACATCAAAAGGATAGATGTAAAAAGACACCAAGGAAGCTTTAGAGAGATA  
 C GC C C G

Inact.4  
 GAGGAAGAGCAAAACAAAGTAAAGAAAAAGGCACAGCAAGCCAGCAGCTGCAGCTGGCACAGGAAACAGCAGCCAGGTC  
 GTCC G C G

AGCCAAAAATTACCCTATAGTGCAGAACCTACAGGGGCCAAATGGTACATCAGGCCATATCACCTAGAACTTTAAATGCA

TGGGTAAAAGTAGTAGAAGAAAGGCTTTTCAGCCCAAGAAATATACCCATGTTTTCAGCATTTATCAGAAAGGAGCCACC

Inact.5  
 CCACAAAGATTAAACACCCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAAATGTTAAAAGAGACTATCAAT  
 G CC G G T G C

GAGGAAGCTCGAGAATGGGATAGAGTGCATCCAGTGCATGCAGGGCCCTATTGCACAGGCCAAATGAGAGAACCAGG

GGAACTGACATAGCAGGAACTACTAGTACCCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTA

Inact.6  
 GGAGAAATCTATAAAGATGGATAATCCTGGGATTAAATAAATAGTAAGAAATGTATAGCCCTACCAGCATTCTGGAC  
 G C G G G C G C G G

ATAAGACAGGACCAAGGAACCCCTTAGAGATTATGTAGACCGGTTCTATAAAACTCTAAGAGCCGGAACAAGCTTCA

FIG. 12A

SUBSTITUTE SHEET (RULE 26)

20 / 131

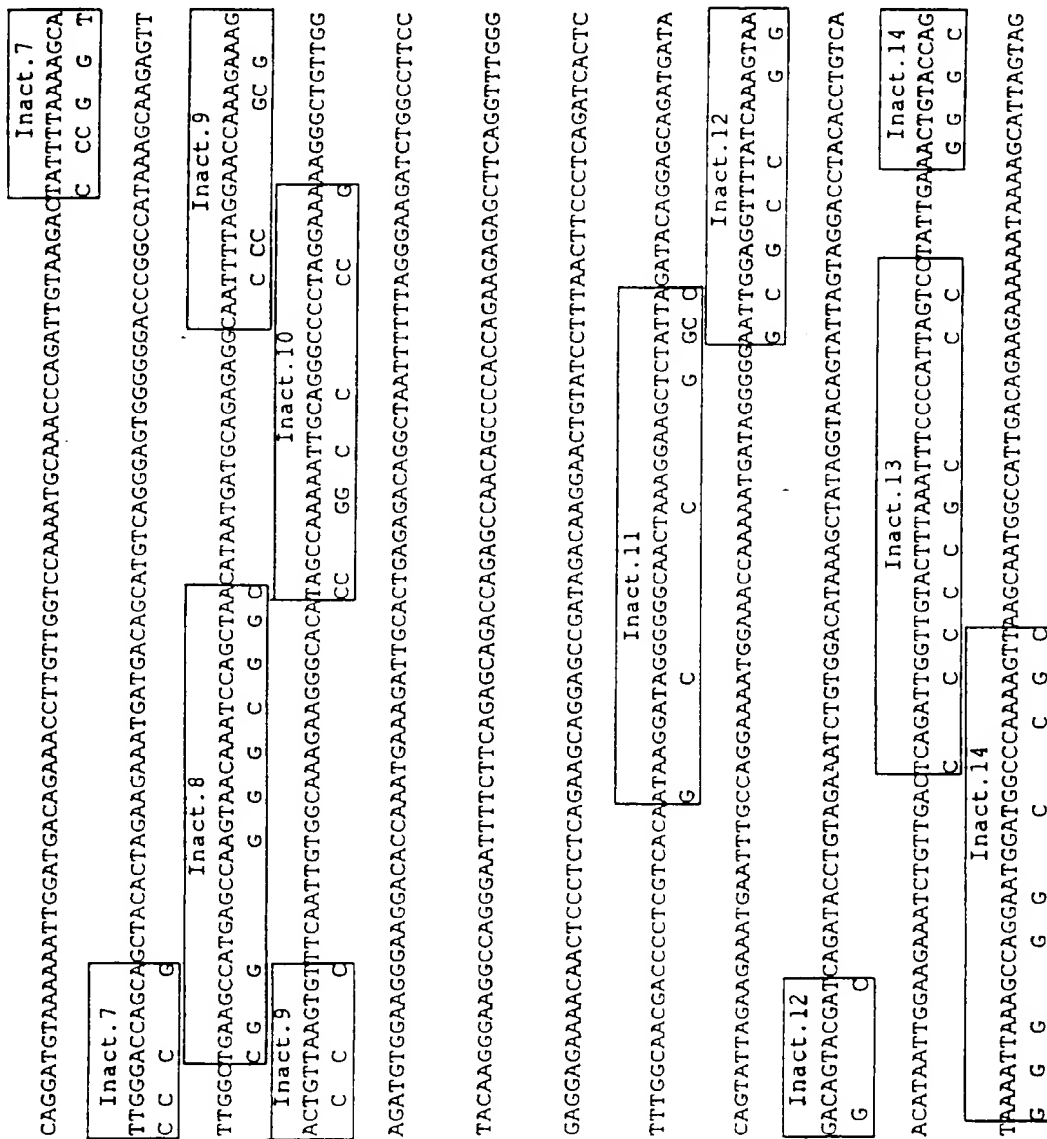


FIG. 12B

SUBSTITUTE SHEET (RULE 26)

21 / 131

AGATATGTACAGAAATGGAAAAAGGAAAAATTTCAAAAATTTGGGCTGAAATCCATACATACTCCAGTATTTG  
CTATAAGAAAAAGACAGTACTAAATGGAGAAAACTAGTAGATTTAGAGAACTTAATAAAAGAACTCAAGACTTCT  
GGGAAGTTGAGTGGAAATACCAACCCCGAGGTTTAAAAAGAGAAAAATCAGTAACAGATTTGGATGGGTGATG  
CATACTTTTCAGTCCCTTAGATAAAGCTTTAGAAAGTATCTGCTTACCATCTAGTATATAAACAATGAGACAC  
CAGGGATTAGATATCAGTACRAATGTCTGCCACAGGGATGGAAAGGATCACCGCAATATTTCCAAAGTAGCATGACAA  
AACTTTAGAGCCCTTTAGAAACAGAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTG  
AATAGAAATAGGGCAGCATAGAAACAAAATAGAGGAACCTGAGACAGCATCTGTTGAGTGGGATTTACCAACACAG  
ACAAAAACATCAGAAAGAACCTCCATTCCTTTGGATGGGTATGAACCTGAGAGTGTGGGAAATTTGAATTTGGGCAAGTCAGA  
TAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGCAATACAGAACTTAAAGCACTAACAGAAATATACCACTAA  
TTTATGAGGGATTAAAGTAAAGCAGTTATGTAACTCTTAGAGGAACCAAGCACTAACAGAAATATATGACCCAT  
CAGAAAGCAGAGCTAGAACTGGCAGAAAAACAGGGAGATCTTAAAGAAACAGCATACATGAAATTTATCAAGAGCCATTTAAAA  
CARAAGACTTAGTAGCAAGAAATACAGAAAGCAGGGCAAGGCCAATGGACATATCAATTTTATCAAGAGCCATTTAAAA  
ATCTGAAAAACAGGAAGTATGCAAGGATGAGGGTGGCCACACTAATGATGTAAAAACAGTTAAACAGAGGCGAGTGCAAA  
AAGTATCCACAGAAAGCATAGTAATATGGGGAAGATTCCTAAATTTAACTACCCCATCAAAAGGAAACATGGGAAG  
CATGGTGGATGGAGTATTTGGCAAGCTACCTGGATTCCTGAGTGGGAGTTTCTATGTAGTGGGCGAGCTAATAGGAGACTAAAT  
GGTACCAAGTTAGAGAAAGAACCCATAGTAGGAGCAGAAACCTTTCTATGTAGTGGGCGAGCTAATAGGAGACTG  
TAGGAAAAACAGGATATGTTACTGACAGAGGAAGACAAAAAGTTGTCTCCATAGCTGACACAAACATCAGAAAGACTG  
AATTACAGCAATTCATCTAGCTTTGCAGGATTCGGGATTAGAGTAACTAAGTAACTAGTAACAGACTCACAATATGCATTAG  
GAATCATTCAAGCACAACCCAGATAAGAGTGAATCAGAGTTAGTCAGTCAAAATTAATAGAGCAGTTAATAAAAAAGGAAA  
AGGCTACCTGGCATGGGTACCAACACACAAAGGAATTTGGAGGAATGAAAGAAACATGAGAAATATCACAGTAATTTGGAGAGCAA  
TCAGGAAAGTACTATTTTGAATGGAATAGATAAGGCCCAAGAAATAGTAGCCAGCTGTGATAAATGTACGTAAAGGAG  
TGGCTAGTATTTAACTGCCACCTGTAGTAGCAAAAGAAATATGGCACTAGATTGTACACATCTAGAGGAAATTTATCC  
AAGCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATAGAGCAATAGAGCAATGGCGGCAAGGCAAGCAATAT  
TGCTAGCAGTTTCATGTAGCCAGTGGATATATAGAGCAAGCAAGTATTTCCAGCAGACAGAGGCGGCAAGCAATAT  
TTCTCTTAAAAATTAGCAGGAAGATGGCCAGTAAAAACAAATACATACAGCAATGGCAGCAATTTCCAGTACTACCG  
TTAAGGCCCTGTGTGGTGGCAGGATCAAGAGGAATTTGGCATTCCTACATCCCAAGTCCCAAGTCAAGGAGTAG  
AATCTATGAATATGAATTAAGGAAATTTAGGACAGGTAGAGATCAGGCTGAACACCTTAAGACAGCAGTACAAA  
TGGCAGTATTCATCCACAATTTTAAAGAAAAGGGGGATTGGGGGATACAGTGCAGGGGAAAGAAATAGTAGACATAA  
TAGCAACAGACATACAACTAAAGAACTACAAAGCAATTTACAAAAATTTCAAAATTTTCGGGTTTATTACAGGGACA  
ACAAAGATCCCTTTGGAAAGGACCCAGCAAGCTTCTCTGAAAGGTGAAGGGCAGTAGTAATCAAGATAATAGTG  
ACATAAAGTAGTGCCCAAGAAAGAAAGCAAAATCATTAGGGATTATGGAAACAGATGGCAGGTGATGATGTGTGG  
CAAGTAGACAGGATGAGGATTAG

FIG. 12C

SUBSTITUTE SHEET (RULE 26)

22 / 131

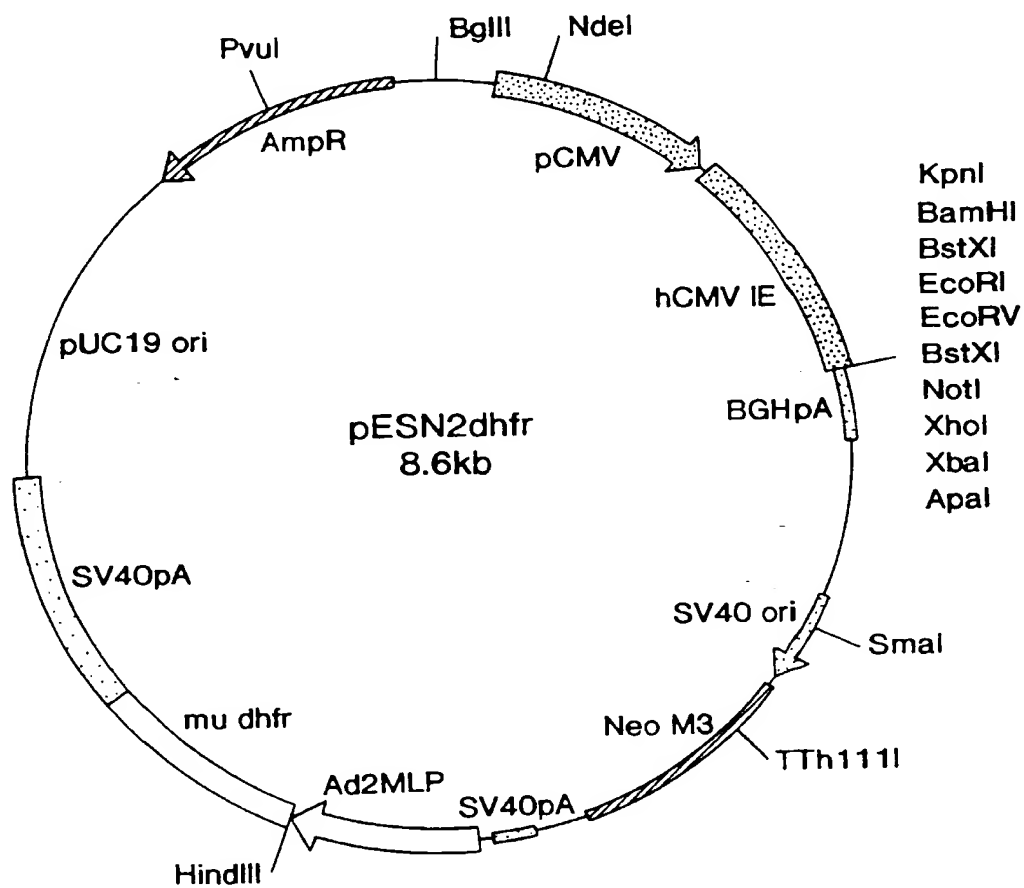


FIG. 13A

SUBSTITUTE SHEET (RULE 26)



23 / 131

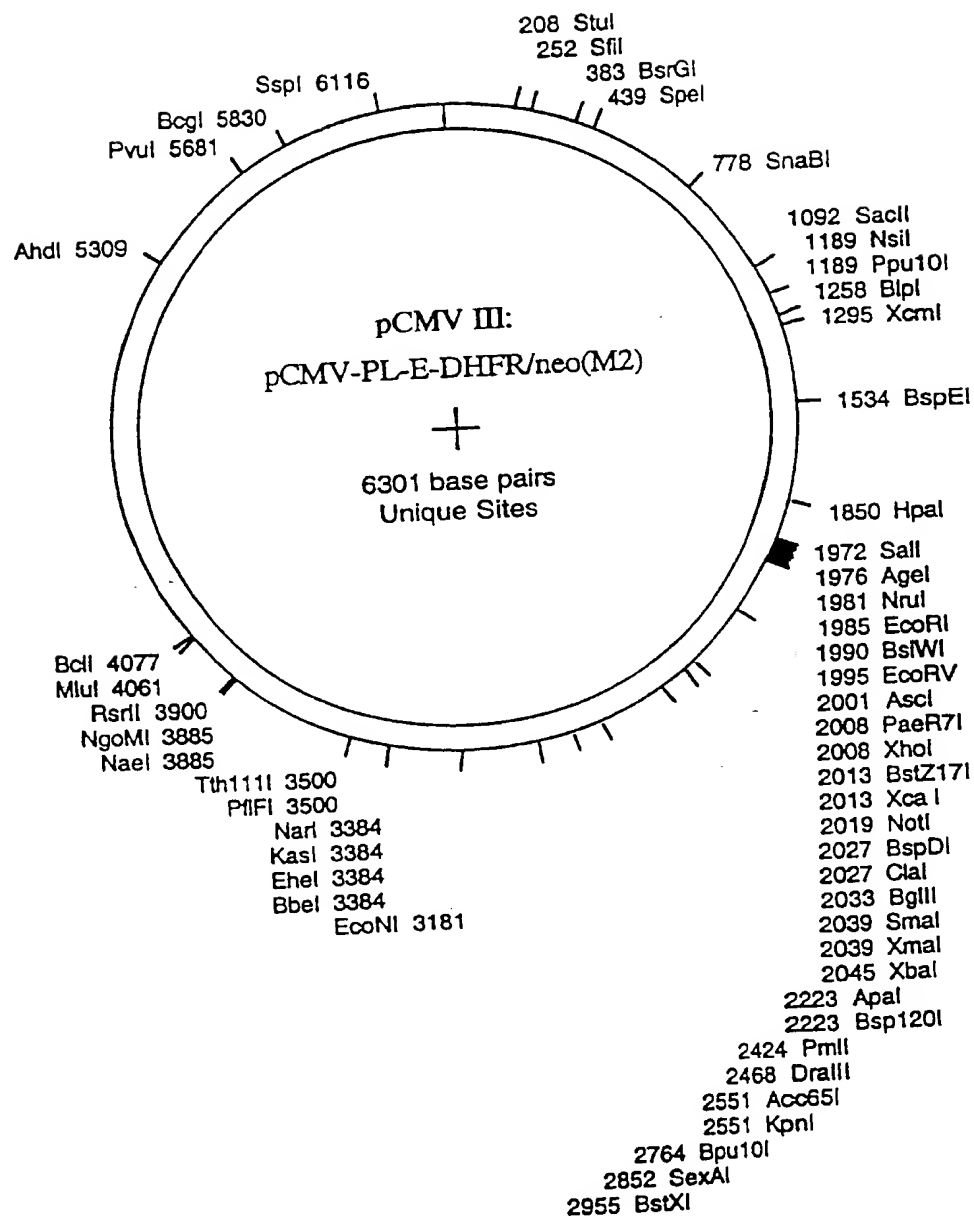


FIG. 13B

SUBSTITUTE SHEET (RULE 26)

24 / 131

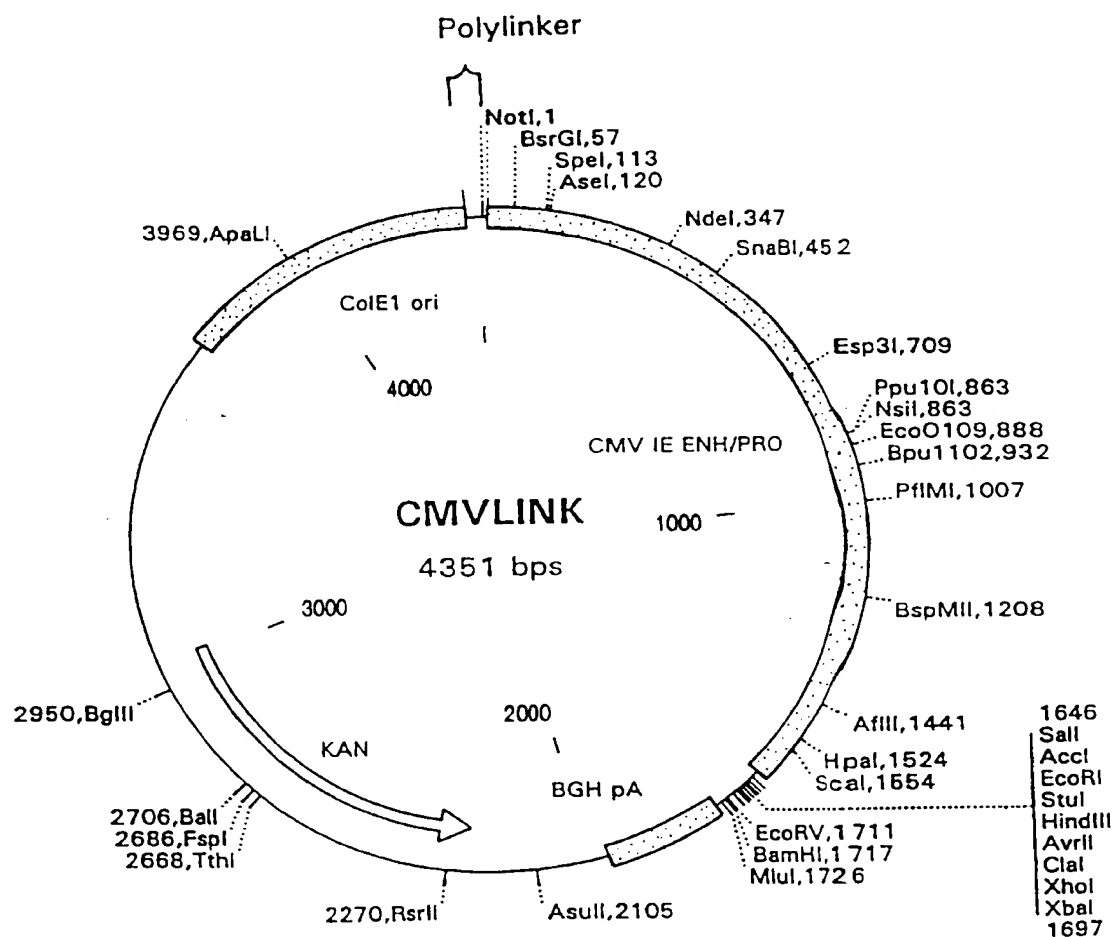


FIG. 14

SUBSTITUTE SHEET (RULE 26)

25 / 131

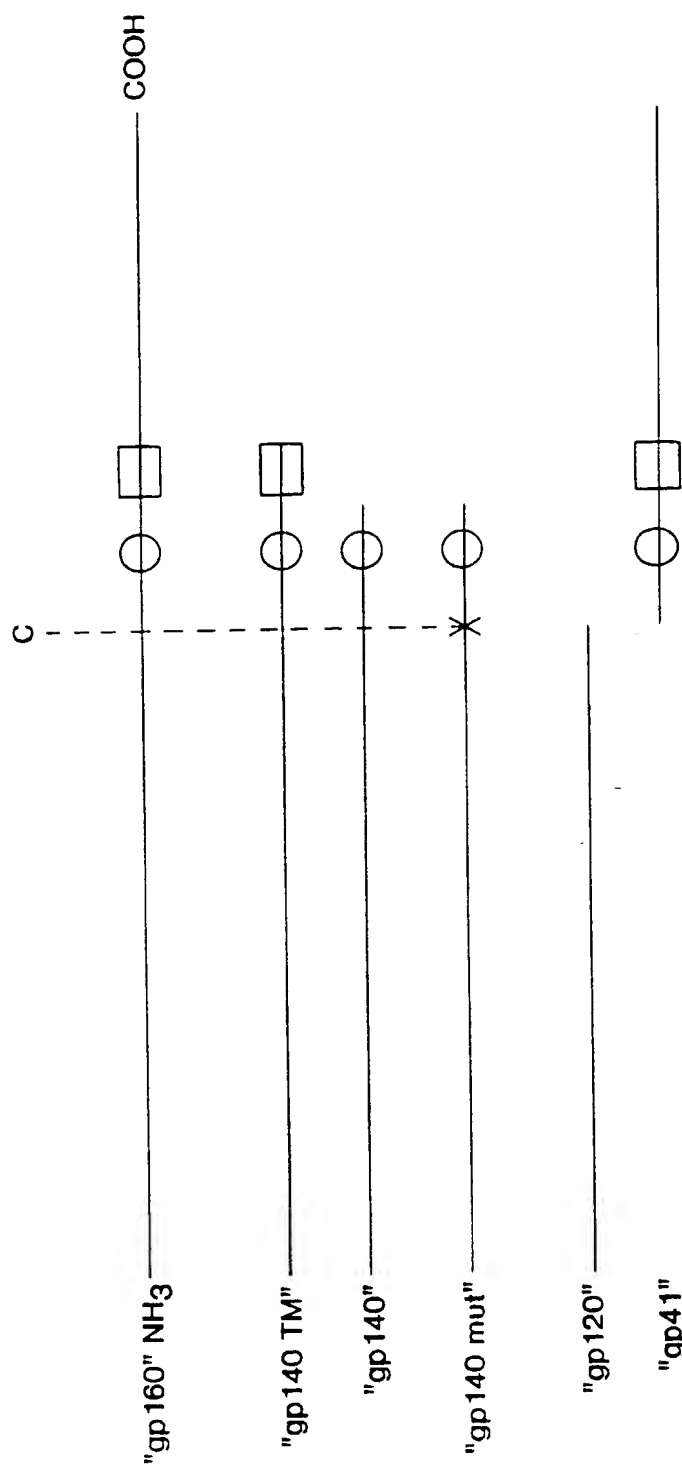


FIG. 15

SUBSTITUTE SHEET (RULE 26)

26 / 131

## gp120wtSF162

GTAGAAAAATTGTGGTCAAGTCTATTATGGGTACCTGTGTGGAAAGCAACCACCCTCTATTTT  
GTGCATCAGATGCTAAGCCTATGACACAGAGGTACATAATGCTGGGCCACACATGCCTGTGTACCCAC  
AGACCTAACCCACAAGAAATAGTATTGGAAATGTGACAGAAAAATTTTAAATGTGGAAAAATAACATG  
GTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAGTCTAAAGCCATGTGTAAAGTTAACCC  
CACTCTGTGTTACTCTACATTGCACCTAATTTGAAGAAATGCTACTAATACCAAGAGTAGTAATTGGAAAGA  
GATGGACAGAGGAGAAATAAAAAATTTGCTCTTTCAAGGTCAACCAAGCATAAGAAATAAGATGCAGAAA  
GAATATGCACCTTTTATAAACTTGATGTAGTACCAATAGATAATGATAATACAAAGCTATAAAATTGATAA  
ATTGTAAACACCTCAGTCATTACACAGGCCCTGTCCAAAGGTATCCTTTGAACCAATTTCCCATACATTATTG  
TGCCCGGCTGGTTTGGGATTTCTAAAGTGTAAATGATAAGAAAGTTCAATGGATCAGGACCATGTACAAAT  
GTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAAATTGCTGTTAAATGGCAGTC  
TAGCAGAAAGAGGGGTAGTAATTAGATCTGAAAAATTTTCACAGACAATGCTAAAACTATAATAGTACAGCT  
GAAGGAATCTGTAGAAATTAATTGTACAAGACCTAACATAATACAAGAAAAAGTATAACTATAGGACCG  
GGGAGAGCATTTTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTTGTAACATTAGTGGAG  
AAAAATGGAATAACACTTTAAACAGATAGTTACAAAATTACAAGCACAAATTTGGGAATAAAACAATAGT  
CTTTAAGCAATCCTCAGAGGGGACCCAGAAAATTGTAATGCACAGTTTAAATTGTGGAGGGAAATTTTTC  
TACTGTAATTCAACACAGCTTTTAAATAGTACTTGGATAATACTATAGGGCCAAATAACACTAATGGAA  
CTATCACACTCCCATGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGGAAAAAGCAATGTATGC  
CCCTCCCATCAGAGGACAAATTAGATGCTCATCAATATTACAGGACTGCTATTAAACAAGAGATGGTGGT  
AAAGAGATCAGTAACACCCCGAGATCTTCAGACCTGGAGGTGGAGATATGAGGGACAAATTGGGAGAAGTG  
AATTATATAAATATAAAGTAGTAAAAAATTGAGCCATTAGGAGTAGCACCCCAAGGCAAGAGAGAGAGT  
GGTGCAGAGAGAAAAAAGA

## FIG. 16

(SEQ ID NO:30)

27 / 131

gp140wtSF162

GTAGAAAAATTGTGGGTACAGTCTATTATGGGTACCTGTGTGGAAGAAGCAACCACCACCTCTATTTT  
GTGCATCAGATGCTAAAGCCTATGACACAGAGGTACATAATGTCTGGGCCACACATGCTGTGTACCCAC  
AGACCCCTAACCCACAAGAAATAGTATTGGAAAATGTGACAGAAAAATTTTAAACATGTGGAATAAATACATG  
GTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGTCTAAAGCCATGTGTAAAGTTAAACCC  
CACTCTGTGTACTCTACATTGCACATAATTTGAAGATGCTACTAATACCAGAGTAGTAATTTGGAAAAGA  
GATGGACAGAGGAGAAATAAAAAATTTGCTCTTTCAAGGTCAACCAAGTCAATAGATAATACAAGCTATAAATTGATAA  
GAATATGCACCTTTTATAAACTTGATGTAGTACCAATAGATAATGATAAATCAATTTCCCATACATTTATTG  
ATTGTAAACACCTCAGTCATTACACAGGCCCTGTCCAAAGGTAATCCTTTGAACCAATTTCCCATACATTTATTG  
TGCCCCGGCTGTTTGGGATTTCTAAAGTGTAAATGATAAAGAAAGTTCAATGGATCAGGACCATGTACAAAT  
GTCAGCACAGTAGAATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAATTTGCTGTTAAATGGCAGTC  
TAGCAGAAAGAGGGTAGTAATTAGATCTGAAAATTTACAGACAAATGCTAAAACTATAATAGTACAGCT  
GAAGGAATCTGTAGAAATTAATTGTACAAGACCTAACATAATACAGAAAAAGTATAACTATAGGACCG  
GGGAGAGCATTTTATGTCAACAGGAGACATAATAGGAGATAAAGACAAGCACATTTGTAACATTAGTGGAG  
AAAAATGGAATAACACTTTAAAAACAGATAGTTACAAAATTACAGCACAATTTGGGAATAAAACAATAGT  
CTTTAAGCAATCCTCAGGAGGGGCCCAAGAAATTTGTAATGACACAGTTTAAATTTGTGGAGGGAAATTTTC  
TACTGTAAATTCACACAGCTTTTAAATAGTACTTGGAAATAATACTATAGGCCAAATAACACTAATGGAA  
CTATCACACTCCCATGCAGAAATAAACAAATTTATAACAGGTGGCAGGAAGTAGGAAAAAGCAATGTATGC  
CCCTCCCATCAGAGGACAAAATTAGATGCTCATCAATAATTACAGGACTGCTATTACAAAGAGATGGTGGT  
AAAGAGATCAGTAACACCAACCGAGATCTTCAGACCTGGAGGTGGAGATATGAGGGACAATTTGGAGAAAGT  
AATTATATAATAAAGTAGTAAAAATTTAGCCATTTAGGAGTAGCACCCACCAAGGCCAAAGAGAGAGT  
GGTGACAGAGAGAAAAAAGAGCAGTGACGCTAGGAGCTATGTTCTTGGGTTCTTGGGAGCAGCAGGAAGC  
ACTATGGGCGCAGGTCAGCTGACGCTGACGGTACAGGCCAGACAATTAATGCTGTTAGTGTGCAACAGC  
AGAACAAATTTGCTGAGAGCTATTGAGGGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCA  
GCTCCAGGCAAGAGTCTGCTGTGGGAAAGATACCTAAAGGATCAACAGCTCCTAGGGATTTGGGGTTGC  
TCTGGAAAACTCATTTGCACCTGCTGTGCTTGGAAATGCTAGTTGGAGTAATAAATCTCTGGATCAGA  
TTTGGGAATAACATGACCTGGATGGAGTGGGAGAGAGAAATTTGACAAATTACACAACTTAATATACACCTT  
AATTGAAGAAATCGCAGAACCAACAAGAAAGAAATGAACAAGAAATTTAGAAATTTGGATAAGTGGGCAAGT  
TTGTGGAATTTGGTTTGACATATCAAAAATGGCTGTGGTATATA

FIG. 17

(SEQ ID NO:31)

SUBSTITUTE SHEET (RULE 26)

28 / 131

gp160wtSF162

GTAGAAAAATTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAGCAACCACCCTCTATTTT  
GTGCATCAGATGCTAAAGCCTATGACACAGAGGTACATAATGTCTGGGCCACACATGCCTGTGTACCCAC  
AGACCCTAACCCACAAGAAATAGTATTGGAAAAATGTGACAGAAAAATTTAACATGTGGAAAAATAACATG  
GTAGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGTCTAAAGCCATGTGTAAAGTTAACCC  
CACTCTGTGTACTCTACATTGCACTAATTTGAAGAATGCTACTAATACCAAGAGTAGTAATTGGAAAGA  
GATGGACAGAGGAGAAATAAAAAATTGCTCTTTCAAGGTCACCACAAGCATAAGAAATAAGATGCAGAAA  
GAATATGCACTTTTTTATAAACTTGATGTAGTACCAATAGATAATGATAATACAAGCTATAAATTGATAA  
ATTGTAACACCTCAGTCATTACACAGGCCTGTCCAAAGGTATCCTTTGAACCAATTCCCATACATTATTG  
TGCCCCGGCTGGTTTTGCGATTCTAAAAGTGTAATGATAAGAAGTTCATGGATCAGGACCATGTACAAAT  
GTCAGCACAGTACAATGTACACATGGAATTAGGCCAGTAGTGTCAACTCAATTGCTGTTAAATGGCAGTC  
TAGCAGAAGAAGGGGTAGTAATTAGATCTGAAAAATTTACAGACAATGCTAAAACCTATAATAGTACAGCT  
GAAGGAATCTGTAGAAATTAATTGTACAAGACCTAAACAATAATACAAGAAAAAGTATAACTATAGGACCG  
GGGAGAGCATTTTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACATTAGTGGAG  
AAAAATGGAATAACACTTTAAACAGATAGTTACAAAATTACAAGCACAAATTTGGGAATAAAACAATAGT  
CTTTAAGCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGGGAATTTTTTC  
TACTGTAATTC AACACAGCTTTTTAATAGTACTTGGAAATAATACTATAGGGCCAAATAACACTAATGGAA  
CTATCACACTCCCATGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGGAAAAGCAATGGTATGC  
CCCTCCCATCAGAGGACAAATTAGATGCTCATCAATATTACAGGACTGCTATTAACAAGAGATGGTGGT  
AAAGGATCAGTAACACCACCGAGATCTTCAGACCTGGAGGTGGAGATATGAGGGACAATTGGAGAAGTG  
AATTATATAAATATAAAGTAGTAAAAATTGAGCCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGT  
GGTGCAGAGAGAAAAAAGAGCAGTGACGCTAGGAGCTATGTTCTTGGGTTCTTGGGAGCAGCAGGAAGC  
ACTATGGGCGCACGGTCACTGACGCTGACGGTACAGGCCAGACAATTATTGTC TGGTATAGTGCAACAGC  
AGAACAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCA  
GCTCCAGGCAAGAGTCTTGGCTGTGGAAAGATACCTAAAGGATCAACAGCTCCTAGGGATTTGGGGTTGC  
TCTGGAAAACCTATTTGCACCACTGCTGTGCCTTGAATGCTAGTTGGAGTAATAAATCTCTGGATCAGA  
TTTGGAATAACATGACCTGGATGGAGTGGGAGAGAGAAATTGACAATTACACAACTTAATATACACCTT  
AATTGAAGAATCGCAGAACCAACAAGAAAAGAAATGAACAAGAATTATTAGAATTGGATAAGTGGGCAAGT  
TTGTGGAATTGGTTTGACATATCAAAATGGCTGTGGTATATAAAAAATATTCATAATGATAGTAGGAGGTT  
TAGTAGGTTTAAAGGATAGTTTTTACTGTGCTTTCTATAGTGAATAGAGTTAGGCAGGGATACTCACCATT  
ATCATTTTCAGACCCGCTTCCAGCCCCAAGGGGACCCGACAGGCCCCAAGGAATCGAAGAAGAAGGTGGA  
GAGAGAGACAGAGACAGATCCAGTCCATTAGTGCATGGATTATTAGCACTCATCTGGGACGATCTACGGA  
GCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTAATCTTGATTGCAGCGAGGATTGTGGAACCTCT  
GGGACGCAGGGGGTGGGAAGCCCTCAAGTATTGGGGGAATCTCCTGCAGTATTGGATTTCAGGAACATAAG  
AATAGTGTCTGTAGTTTGTGTTGATGCCATAGCTATAGCAGTAGCTGAGGGGACAGATAGGATTATAGAAG  
TAGCACAAAGAATTGGTAGAGCTTTTCTCCACATACCTAGAAGAATAAGACAGGGCTTTGAAAGGGCTTT  
GCTATAA

FIG. 18

(SEQ ID NO:32)

SUBSTITUTE SHEET (RULE 26)

gp120.modSF162

gaatctgcaccatggatgcaatgaagagagggctctgctgtgtgctgtgtgtgagcagtc  
tctcgttctgcaccagcctggagaagctgtgggtgacctgtactacggcgtgccctgtggaag  
gaggccaccaccacctgtctgcgcagcgacgccaaggcctacgacaccgaggtgcacaactg  
tggccaccaccgctgcgtgccaccgaccccaaccccccaggagatcgtgctggagaactgacc  
gagaaactcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagcctgcgtgaagctgacccccctgtgcgtgacctgcaccaacctg  
aagaacgccaccaacaacagagcagcaactggaaggagatggaccggcgagatcaagaactgc  
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccctgttctacaagctg  
gacgtggtgccatcgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
accagggcctgccccaggctgagcttcgagcccatccccatccactactgcgccccgcggcttc  
gccatcctgaagtgcacgacaagaagtacaaggcagcgccccctgcaccaactgagcacctg  
cagtgacccaccggcatccgcccctgggtgagcaccacctgctgctgaacggcagcctggccgag  
gaggggctggtgatccgcagcgagaacttcaccgacaacgcccaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcaccggcccccaacaacaccccgcaagagcatcaccatcggcccc  
ggcgcgccttctacgcccgcgacatcatcggcgacatccgccaggccccactgcaacatcagc  
ggcgagaagtggaacaacacctgaagcagatcgtgaccaagctgcaggccccagctcggcaacaag  
accatcgtgttcaagcagagcagcgggcgaccccgagatcgtgatgcacagcttcaactgcggc  
ggcgagtcttctactgcaacagcaccagctgttcaacagcacctggaacaacaccatcggcccc  
aacaacaccaaggcaccatcacctgccctgccgatcaagcagatcatcaaccgtggcaggag  
gtgggcaaggccatgtacgcccccccatccggggccagatccgctgcagcagcaacatcacggc  
ctgctgctgacccgcagcggcggaaggagatcagcaacaccaccagatcttccgccccggcggc  
ggcgacatgcgcgaactggcgagcgagctgtacaagtacaagtggtgaagatcgagccccctg  
ggcgtggcccccaacgaaggccaagcccgctgggtgcagcgcgagaagcgttaactcgag

**FIG. 19**  
(SEQ ID NO:33)





31 / 131

gp120.modsF162.delV1V2

[illegible]

**FIG. 21**  
(SEQ ID NO:35)

32 / 131

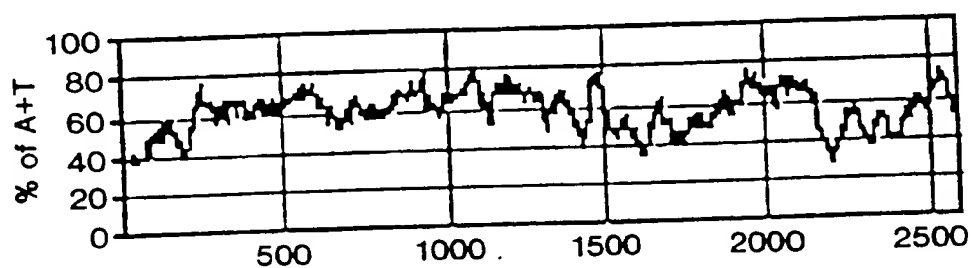


FIG. 22A

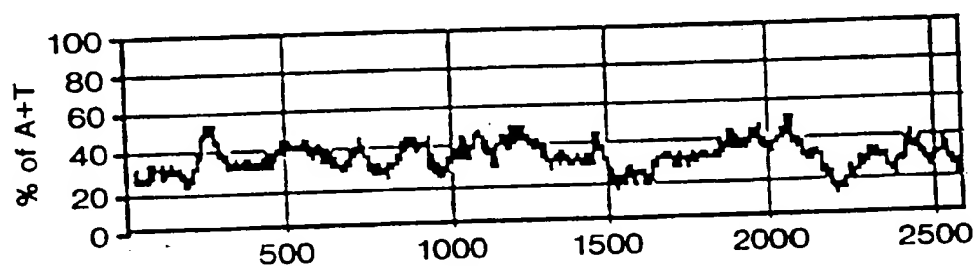


FIG. 22B

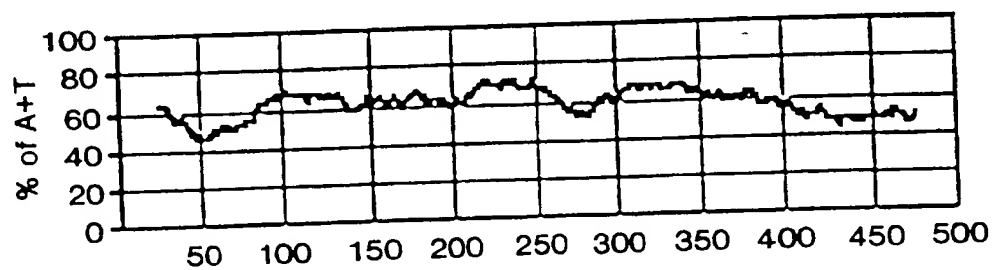


FIG. 22C

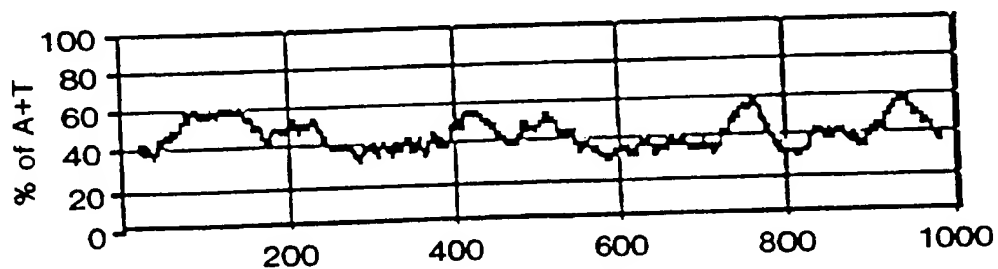


FIG. 22D

SUBSTITUTE SHEET (RULE 26)

33 / 131

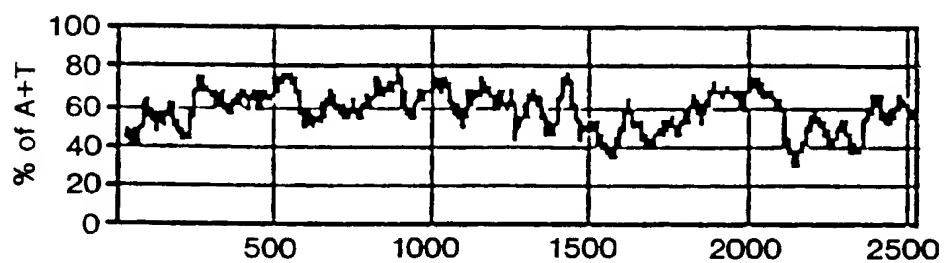


FIG. 22E

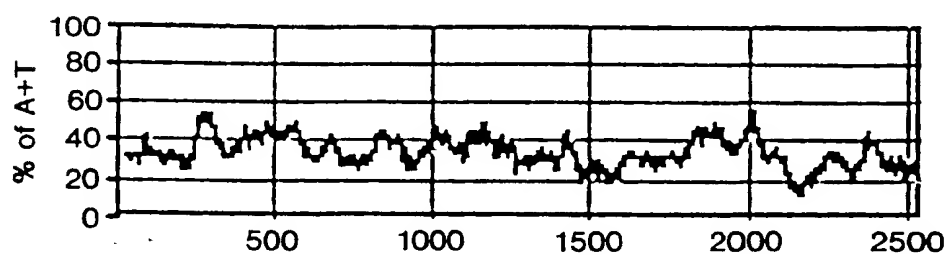


FIG. 22F

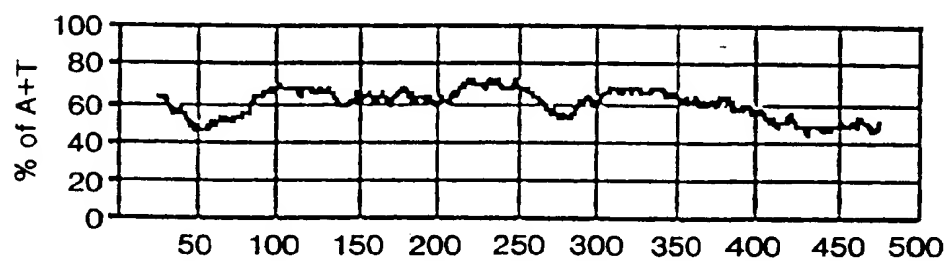


FIG. 22G

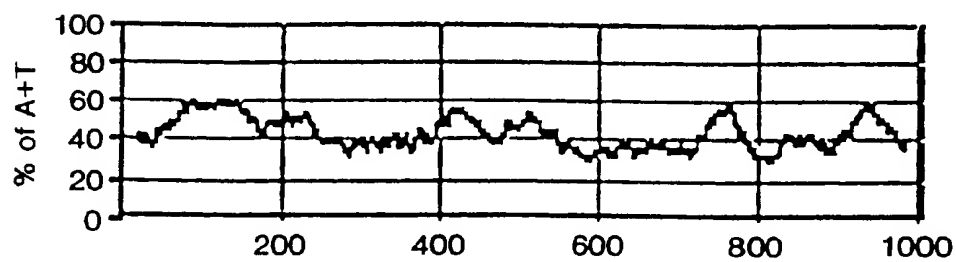


FIG. 22H

SUBSTITUTE SHEET (RULE 26)

34 / 131

## gp140.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcggtgcccgtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggccaccacgcctgctgccccaccgacccccacccccaggagatcgtgctgggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcaactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc  
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccctgttctacaagctg  
gacgtgggtgccccatcgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
accagggcctgcccccaaggtgagcttcgagccccatccccatccactactgcgcccccgccggttc  
ggccatcctgaagtgcacgacaagaagttcaacggcgagcgccccctgcaccaacgtgagcaccgtg  
cagtgcacccacggcatccgccccgtgggtgagcaccacagctgctgctgaacggcgagcctggccgag  
gagggcggtggtgatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcacccgcccccaacaacaacacccgcaagagcatcaccatcggtcccc  
ggccgcgccttctacgccaccggcgacatcatcgggcgacatccgccaggccccactgcaacatcagc  
ggcgagaagtggacaacaccctgaagcagatcgtgaccaagctgcaggccccagttcggaacaag  
accatcgtgttcaagcagagcagcgggcgagcccccgagatcgtgatgcacagcttcaactgcggc  
ggcgagttcttctactgcaacagcaccacagctgttcaacagcacctggaacaacaccatcggtccc  
aacaacaccaacggcaccatcacctgccccgcccgcagatcaagcagatcatcaaccgctggcaggag  
gtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaacatcacccggc  
ctgctgctgacccgcgacggcggaaggagatcagcaacaccaccagatcttccgccccggcggc  
ggcgacatgcgcgacaactggcgagcgagctgtacaagtacaagggtggtgaagatcgagccccctg  
ggcggtggccccccaccaaggccaagcgccgcgtggtgcagcgcgagaagcgcgccgtgaccctgggc  
gccatgttccctgggcttccctggggcgccgcggcgagcaccatggggcgcccgagcctgaccctgacc  
gtgcaggccccgcagctgctgagcggcacgtgcagcagcagaacaacctgctgcgcgccatcgag  
gcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggccccgcgtgctggcc  
gtggagcgctacctgaaggaccagcagctgctggggcatctggggctgcagcggcaagctgatctgc  
accaccgcccgtgcccctggaacgccagctggagcaacaagagcctggaccagatctggaacaacatg  
acctggatggagtgaggagcgagatcgacaactacaccaacctgatctacaccctgatcgaggag  
agccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtgg  
aactggttcgacatcagcaagtggctgtggtactctaactcgag

**FIG. 23**  
(SEQ ID NO:36)

*SUBSTITUTE SHEET (RULE 26)*

35 / 131

gp140.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggcaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcccgcgagatcaagaactgc  
agcttcaagggtgggcccggcaagctgatcaactgcaacaccagcgtgatcaccagggcctgcccc  
aagggtgagcttcgagcccatcccatccactactgcgcccccgccggcttcgccatcctgaagtgc  
aacgacaagaagttcaacggcagcggccccctgcaccaacgtgagcaccgtgcagtgcaccacggc  
atccgccccctgggtgagcaccagctgctgctgaacggcagcctggccgaggagggcgtgggtgatc  
cgcagcggagaacttcaccgacaacgcccaagaccatcatcgtgcagctgaaggagagcgtggagatc  
aactgcaccggccccaacaacaacaccggcaagagcatcaccatcgcccccgccgcgccttctac  
gccaccggcgacatcatcgccgacatccgccaggcccaactgcaacatcagcggcgagaagtggaac  
aacaccctgaagcagatcgtgaccaagctgcaggcccaagtccggcaacaagaccatcgtgttcaag  
cagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggcgccgagttcttctac  
tgcaacagcaccagctgttcaacagcacctggaacaacaccatcgcccccaacaacaccaacggc  
accatcacccctgcccctgcccgcacatcaagcagatcatcaaccgctggcaggaggtgggcaaggccatg  
tacgcccccccccatccgcccgcagatccgctgcagcagcaacatcacccgcccctgctgctgacccgc  
gacggcgccgaaggagatcagcaacaccaccgagatcttccgccccggcgccggcgacatgcgcgac  
aactggcgccagcagctgtacaagtacaaggtgggtgaagatcgagccccctgggctggccccacc  
aaggccaagcgcgcgtgggtgcagcgcgagaagcgcgcctgacccctgggcccacatgttcctgggc  
ttcctggggcgccgcggcagcaccatgggcccgcagcctgacccctgacccgtgcaggccccgcag  
ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgaggcccagcagcacctg  
ctgcagctgaccgtgtggggcatcaagcagctgcaggccccgcgtgctggccgtggagcgtacctg  
aaggaccagcagctgctgggcatctggggctgcagcggcaagctgatctgcaccaccgcccgtgccc  
tggaacgccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtgg  
gagcgcgagatcgacaactacaccaacctgatctacaccctgatcgaggagagccagaaccagcag  
gagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtggaactgggttcgacatc  
agcaagtggctgtggtacatctaactcgag

**FIG. 24**

(SEQ ID NO:37)

*SUBSTITUTE SHEET (RULE 26)*

36 / 131

gp140.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcggtgcccggtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgccccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctggggcgccggcaactgccagacc  
agcgtgatcaccagggcctgccccaaagggtgagcttcgagccccatccccatccactactgcgcccc  
gcccggcttcgccatcctgaagtgcacgcacaagaagtccaacggcagcgggccccctgcaccaacgtg  
agcaccgtgcagtgcacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagc  
ctggccgaggagggcggtggtgatccgcagcgagaacttcaccgacaacggccaagaccatcatcggtg  
cagctgaaggagagcggtggagatcaactgcacccgcccccaacaacaacaccccgcaagagcatcacc  
atcgggccccggcgcgcttctacgcccacggcgacatcatcgggcagatccgcccaggcccactgc  
aacatcagcgggcgagaagtggacaacacccctgaagcagatcgtgaccaagctgcaggcccagttc  
ggcaacaagaccatcgtgttcaagcagagcagcgggcgagccccgagatcgtgatgcacagcttc  
aactgcggcgggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacacc  
atcgggcccaacaacaccaacggcaccatcacccctgcccctgcccgcacatcaagcagatcatcaaccgc  
tggcaggaggtgggcaaggccatgtacgccccccccatccgcgggccagatccgctgcagcagcaac  
atcacccggcctgctgctgacccgcgacggcgggcaaggagatcagcaacacaccagagatcttccgc  
ccccggcgggcgacatgcgcgacaactggcgcgagcagctgtacaagtaacagggtggtgaagatc  
gagccccctggcggtggccccccaccaaggccaagcgcccgctggtgcagcgcgagaagcgcgccgtg  
accctggggcgccatgttcttgggcttcttggggcgccggcagcaccatgggcgccccgagcctg  
accctgacccgtgcaggcccccgagctgctgagcgccatcgtgcagcagcagaacaacctgctgcgc  
gccatcgaggcccagcagcacctgctgcagctgacccgtgtggggcatcaagcagctgcaggcccgc  
gtgctggccgtggagcgtacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag  
ctgatctgcaccacgcctgcccctggaacgccagctggagcaacaagagcctggaccagatctgg  
aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacaccctg  
atcgaggagagccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggccc  
agcctgtggaactgggttcgacatcagcaagtggctgtggtacatctaactcgag

**FIG. 25**

(SEQ ID NO:38)

*SUBSTITUTE SHEET (RULE 26)*

37 / 131

gp140.mut.modSF162

gaattcgccaccatggatgcaatgaagagaggggtctgctgtgtgctgctgctgtgtgtggagcagtc  
ttcggtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggcaccacgcctgctgccccaccgacccccacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcactgcaccaacctg  
aagaacgcccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc  
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccctgttctacaagctg  
gacgtggtgcccacgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
accaggcctgccccaaaggtgagcttcgagccccatccccatccactactgcgcccccgccggcttc  
gccatcctgaagtgcacgacaagaagtccaacggcagcgccccctgcaccaacgtgagcaccgtg  
cagtgcacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagcctggccgag  
gaggcggtggtgatccgacgcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcacccgccccaaacaacacccgcaagagcatcaccatcgcccc  
ggcgcgccttctacgccaccggcgacatcatcgccgacatccgcccaggcccaactgcaacatcagc  
ggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggccccagttcggcaacaag  
accatcgtgttcaagcagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcccc  
ggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacaccatcgcccc  
aacaacaccaacggcaccatcacccctgcccctgcccgcataagcagatcatcaaccgctggcaggag  
gtgggcaaggccatgtacgccccccccatccgcccgcagatccgctgcagcagcaacatcacccggc  
ctgctgctgacccgcgacggcggaaggagatcagcaacaccaaccgagatcttccgccccggcggc  
ggcgacatgcgcgacaactggcgcgagcagctgtacaagtacaagggtggtgaagatcgagccccctg  
ggcgtggccccccaccaaggccaagcgcccgctggtgcagcgcgagaagagcgccgtgacccctgggc  
gccatgttccctgggcttccctgggcgcccgccgacacccatggcgccccgcagcctgacccctgacc  
gtgcaggccccgcagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgag  
gcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggccccgcgtgctggcc  
gtggagcgtacctgaaggaccagcagctgctgggcctctggggctgcagcggcaagctgatctgc  
accaccgcccgtgccctggaacgccagctggagcaacaagagcctggaccagatctggaacaacatg  
acctggatggagtgaggagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggag  
agccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtgg  
aactggttcgacatcagcaagtggctgtggtacatctaaactcgag

FIG. 26

(SEQ ID NO:39)

SUBSTITUTE SHEET (RULE 26)

38 / 131

gp140.mut.modSF162.delV2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcggtgcccgtgtggaag  
gagggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc  
agcttcaaggtggggcgccggaagctgatcaactgcaacaccagcgtgatcaccagggcctgcccc  
aaggtgagcttcgagcccatcccatccactactgccccccgcccggcttcgccatcctgaagtgc  
aacgacaagaagttcaacggcagcgccccctgcaccaacgtgagcacctgacgtgcacccacggc  
atccgccccgtggtgagcaccacagctgctgctgaacggcagcctggccgaggagggcggtggtgatc  
cgcagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc  
aactgcacccgcceccaacaacaacacccgcaagagcatcaccatcgggcccggccgcgccttctac  
gccaccggcgacatcatcggcgacatccgccaggcccactgcaacatcagcggcgagaagtggaaac  
aacaccctgaagcagatcgtgaccaagctgcaggccccagttcggcaacaagaccatcgtgttcaag  
cagagcagcgggcgggcgaccccgagatcgtgacacagcttcaactgcggcgggcgagttcttctac  
tgcaacagcaccacagctgttcaacagcacctggaacaacaccatcgggcccacaacaccaacggc  
accatcacccctgcccctgcccgcacatcaagcagatcatcaaccgctggcaggaggtgggcaaggccatg  
tacgccccccccatccgcgggccagatccgctgcagcagcaacatcacggccctgctgctgacccgc  
gacggcgggcaaggagatcagcaacaccaccgagatcttccgccccggcgggcgacatgcgcgac  
aactggcgcgagcagctgtacaagtacaagggtggtgaagatcgagcccctgggctggccccccacc  
aaggccaagcgccgctggtgacgcgcgagaagagcgccgtgaccctgggcgccatgttcttgggc  
ttctctggggcgccgcccgcagcaccatggggcgcccgcagcctgaccctgacgctgcaggcccgcag  
ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgaggcccagcagcacctg  
ctgcagctgacccgtgtggggcatcaagcagctgcaggcccgcgtgctggcctgaggcgctacctg  
aaggaccagcagctgctggggcatctggggctgcagcggaagctgatctgcaccaccgcccgtggcc  
tggaacgccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtgg  
gagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggagagccagaaccagcag  
gagaagaacgagcaggagctgctggagctggacaagtggggccagcctgtggaactgggttcgacatc  
agcaagtggctgtggtacatctaactcgag

**FIG. 27**  
(SEQ ID NO:40)

*SUBSTITUTE SHEET (RULE 26)*



39/131

gp140.mut.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccggtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctggtggcgccggcaactgccagacc  
agcgtgatcaccacaggcctgcccccaaggtgagcttcgagcccatcccatccactactgcccccc  
gccccgcttcgccatcctgaagtgcacgacaagaagttcaacggcagcgccctgcaccaacgtg  
agcaccgtgcagtgacccacggcatccgccccgtgggtgagcaccagctgctgctgaacggcagc  
ctggccgaggagggcggtggtgatccgagcgagaacttcaccgacaacgccaagaccatcatcgtg  
cagctgaaggagagcggtggagatcaactgcacccgcccccaacaacaacacccgcaagagcatcacc  
atcgggccccggcgcgcttctacgccaccggcgacatcatcgggcgacatccgccaggcccactgc  
aacatcagcggcgagaaagtgaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttc  
ggcaacaagaccatcgtgttcaagcagagcagcgggcgagccccgagatcgtgatgcacagcttc  
aactgcggcgggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacacc  
atcgggccccaacaacaccaacggcaccatcacctgccccgccccatcaagcagatcatcaaccgc  
tggcaggaggtgggcaaggccatgtacgccccccccatccgccccagatccgctgcagcagcaac  
atcacccggcctgctgctgacccgagcggcggaaggagatcagcaacaccaccagagatcttcgc  
cccgccggcgccgacatgcgcgacaactggcgagcgagctgtacaagtacaaggtggtgaagatc  
gagccccctggcggtggccccccaccaaggccaaagcgccgctggtgcagcgcgagaagagcgccgtg  
accctggggcgccatgttcctgggcttcctggggcgccggcgagcaccatggggcgcccgagcctg  
accctgaccgtgcaggccccagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgc  
gccatcgaggcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggcccg  
gtgctggccgtggagcgctacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag  
ctgatctgcaccaccgctgcccgtggaacgccagctggagcaacaagagcctggaccagatctgg  
aacaacatgacctggatggagtgggagcgagatcgacaactacaccaacctgatctacacctg  
atcgaggagagccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggccc  
agcctgtggaactggttcgacatcagcaagtggctgtggtacatctaactcgag

FIG. 28

(SEQ ID NO:41)

SUBSTITUTE SHEET (RULE 26)

40/131

gp140.mut7.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgacctgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgcccaccgacccccacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcgccgagatcaagaactgc  
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcctgttctacaagctg  
gacgtggtgcccacgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
acccaggcctgccccaaaggtgagcttcgagcccatccccatccactactgcgcccccgccggcttc  
gacctcctgaagtgcacgacaagaagtccaacggcagcgccccctgcaccaacgtgagcaccgtg  
cagtgacccacggcatccgccccgtggtgagcaccagctgctgctgaaaggcagcctggccgag  
gagggcggtggtgatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcacccgccccaaacaacacccgcaagagcatcaccatcgcccc  
ggccgcgcttctacgccaccggcgacatcatcgggacatccgccaggcccaactgcaacatcagc  
ggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggcccaagtccggcaacaag  
accatcgtgttcaagcagagcagcgggcgaccccgagatcgtgatgcacagcttcaactgcggc  
ggcgagtcttctactgcaacagcaccagctgttcaacagcacctggaaacaacaccatcgcccc  
aacaacaccaacggcaccatcacctgccccgcatcaagcagatcatcaaccgctggcaggag  
gtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaacatcacgggc  
ctgctgctgacccgcgacggcggaaggagatcagcaacaccaccagatcttcgccccggcggc  
ggcgacatgcgcgacaactggcgagcgagctgtacaagtacaagggtggtgaagatcgagccccctg  
ggcggtggccccaccaggccatcagcagcgtggtgcagagcgagaagagcgccgtgacctgggc  
gccatgttccctgggcttccctgggcgcccggcgagcaccatgggcgcccgcagcctgacctgacc  
gtgcaggccccgcagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgag  
gcccagcagcacctgctgcagctgacctgtggggcatcaagcagctgcaggccccgctgctggcc  
gtggagcgctacctgaaggaccagcagctgctggggcatctggggctgcagcggcaagctgatctgc  
accaccgcccgtgccctggaacgccagctggagcaacaagagcctggaccagatctggaacaacatg  
acctggatggagtgggagcgcgagatcgacaactacaccaacctgatccacacctgatcgaggag  
agccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtgg  
aactggttcgacatcagcaagtggctgtggtacatctaactcgag

**FIG. 29**

(SEQ ID NO:42)

*SUBSTITUTE SHEET (RULE 26)*

41 / 131

gp140.mut7.modSF162.delV2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgactgcaccaacctg  
aagaacgcccaccaacaccaagagcagcaactggaaggagatggaccgcgcgagatcaagaactgc  
agcttcaaggtgggcgccggaagctgatcaactgcaacaccagcgtgatcaccagggcctgcccc  
aaggtgagcttcgagcccatccccatccactactgcgcccccgccggttcgccatcctgaagtgc  
aacgacaagaagttcaacggcagcgcccccctgcaccaacgtgagcaccgtgcagtgcaccacggc  
atccgccccgtggtgagcaccagctgctgctgaacggcagcctggccgaggaggcggtggtgatc  
cgagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc  
aactgcacccgcccccaacaacaacaccccgcaagagcatcaccatcgcccccgccgccccttctac  
gccaccggcgacatcatcggcgacatccgccaggcccaactgcaacatcagcggcgagaagtggaaac  
aacaccctgaagcagatcgtgaccaagctgcaggcccaagtccggcaacaagaccatcgtgttcaag  
cagagcagcgccgagaccccgagatcgtgatgcacagcttcaactgcggcgggcgagttcttctac  
tgcaacagcaccagctgttcaacagcacctggaacaacaccatcgcccccaacaacaccaacggc  
accatcacccctgcctgcccgcagatcaagcagatcatcaaccgctggcaggaggtgggcaaggccatg  
tacgcccccccccatccgcccgcagatccgctgcagcagcaacatcacggcctgctgctgacccgc  
gacggcggaaggagatcagcaacaccacccagatcttccgccccggcgggcgacatgcgcgac  
aactggcgcgagcgtgtacaagtacaaggtggtgaagatcgagccccctgggctggccccccacc  
aaggccatcagcagcgtggtgcagagcgagaagagcgccgtgaccctgggcgccatgttcttgggc  
ttcttgggcccgcggcagcaccatgggcccgcagcctgaccctgaccgtgcaggcccgcag  
ctgctgagcgccatcgtgcagcagcagaacaacctgctgcgcgccatcgaggcccagcagcacctg  
ctgcagctgaccgtgtggggcatcaagcagctgcaggcccgcgtgctggccgtggagcgtacctg  
aaggaccagcagctgctgggcatctggggctgcagcggcaagctgatctgcaccaccgcccgtgccc  
tggaacgcccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtgg  
gagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggagagccagaaccagcag  
gagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtggaactggttcgacatc  
agcaagtggctgtggtacatctaactcgag

FIG. 30

(SEQ ID NO:43)

SUBSTITUTE SHEET (RULE 26)

42 / 131

gp140.mut7.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gagggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggcaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctggtggcgccggcaactgccagacc  
agcgtgatcaccaggcctgcccccaagggtgagcttcgagcccatcccatccactactgcgcccc  
gccggcttcgccatcctgaagtgaacgacaagaagttaacggcagcgccccctgcaccaacgtg  
agcaccggtgcagtgcacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagc  
ctggccgaggagggcggtggtgatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtg  
cagctgaaggagagcgtggagatcaactgcacccgcccccaacaacaacacccgcaagagcatcacc  
atcgggcccgccgcttctacgcccaccggcgacatcatcggcgacatccgccaggcccactgc  
aacatcagcgccgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggcccagttc  
ggcaacaagaccatcgtgttcaagcagagcagcgccggcgaccccgagatcgtgatgcacagcttc  
aactgcggcgccgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacacc  
atcgggccccaacaacaccaacggcaccatcacctgccccctgcccgcacatcaagcagatcatcaaccgc  
tggcaggaggtgggcaaggccatgtacgccccccccatccgcccagatccgctgcagcagcaac  
atcacccggcctgctgctgacccgcgacggcggaaggagatcagcaacaccaccgagatcttcgc  
cccggcgccgagacatgcgcgacaactggcgagcgagctgtacaagtacaagggtggtgaagatc  
gagccccctggcgctggccccccaccaaggccatcagcagcgtggtgcagagcgagaagagcgccgtg  
accctggggcgccatgttcctgggcttcctggggcgccggcgagcaccatggggcgcccgagcctg  
accctgaccgtgcaggccccccagctgctgagcgccatcgtgcagcagcagaacaacctgctgcgc  
gccatcagggcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggcccgc  
gtgctggccgtggagcgctacctgaaggaccagcagctgctggggcatctggggctgcagcggaag  
ctgatctgcaccaccgcccgtgcccgggaacgcccagctggagcaacaagagcctggaccagatctgg  
aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctg  
atcgaggagagccagaaccagcaggagaagaacgagcaggagctgctcgagctggacaagtgggccc  
agcctgtggaactggttcgacatcagcaagtggctgtggtacatctaactcgag

FIG. 31  
(SEQ ID NO:44)

SUBSTITUTE SHEET (RULE 26)

43 / 131

gpl40.mut8.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcggtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcggtgcccggtgtggaag  
gaggccaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggccaaccacgcctgctgccccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgaccctgcactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcgcgagatcaagaactgc  
agcttcaaggtgaccaccagcatccgcaacaagatgcagaaggagtacgcccctgttctacaagctg  
cagctgggtgccccatcgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
accagggcctgcccccaaggtgagcttcgagcccatccccatccactactgcccccccgccggttc  
gccatcctgaagtgaacgacaagaagttcaacggcagcgccccctgcaccaacgtgagcaccgtg  
cagtgcacccacggcatccgccccgtgggtgagcaccagctgctgctgaaacggcagcctggccgag  
gagggcggtggtgatccgcagcgagaacttcaccgacaacgcgaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcaccgcccccaacaacacccgcaagagcatcaccatcgcccc  
ggccgcgcccttctacgccaccggcgacatcatcgggcagatccgcccaggcccactgcaacatcagc  
ggcgagaagtggaaacaacaccctgaagcagatcgtgaccaagctgcaggcccagttcggaacaag  
accatcgtgttcaagcagagcagcgggcgacccccagatcgtgatgcacagcttcaactgcggc  
ggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaaacaacaccatcgcccc  
aacaacaccaacggcaccatcacctgccccgcatcaagcagatcatcaaccgctggcaggag  
gtgggcaaggccatgtacgccccccccatccgcgccagatccgctgcagcagcaacatcacgggc  
ctgctgctgaccccgacggcggaaggagatcagcaaacaccaccagatcttccgccccggcggc  
ggcgacatgcgcgacaactggcgagcagctgtacaagtacaaggtggtgaagatcgagccctg  
ggcggtggccccaccatcgccatcagcagcgtggtgcagagcgagaagagcgccgtgacctgggc  
gccatgttccctgggcttccctggggcgccggcgagcaccatggggcgcccgagcctgacctgacc  
gtgcaggccccgcagctgctgagcgccatcgtgcagcagcagaacaacctgctgcgccccatcgag  
gcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggccccgctgctggcc  
gtggagcgctacctgaaggaccagcagctgctgggcatctggggctgcagcggaagctgatctgc  
accaccgcccgtgccctggaacgccagctggagcaacaagagcctggaccagatctggaacaacatg  
acctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggag  
agccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtgg  
aactgggttcgacatcagcaagtggctgtggtacatctaactcgag

FIG. 32

(SEQ ID NO:45)

SUBSTITUTE SHEET (RULE 26)

44 / 131

gp140.mut8.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgcctggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcccagcgcagccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgcaactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcgccgagatcaagaactgc  
agcttcaaggtgggcgccggcaagctgatcaactgcaacaccagcgtgatcaccaggcctgcccc  
aaggtgagcttcgagcccatcccatccactactgcgcccccgccggcttcgccatcctgaagtgc  
aacgacaagaagttcaacggcagcggccccctgcaccaacgtgagcaccgtgcagtgcaaccaaggc  
atccgccccgtgggtgagcaccagctgctgctgaacggcagcctggccgaggaggcggtggtgatc  
cgcagcggagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc  
aactgcacccgcccccaacaacaacaccgcgaagagcatcaccatcgcccccgccgcgccttctac  
gccaccggcgacatcatcgggcgacatccgcccaggccactgcaacatcagcggcgagaagtggaaac  
aacaccctgaagcagatcgtgaccaagctgcaggccccagttcggaacaagaccatcgtgttcaag  
cagagcagcggcgccgaccccgagatcgtgatgcacagcttcaactgcggcgccgagttcttctac  
tgcaacagcaccagctgttcaacagcacctggaacaacaccatcgcccccaacaacaccaacggc  
accatcacccctgccctgcgcgatcaagcagatcatcaaccgctggcaggaggtgggcaaggccatg  
tacgcccccccccatccgcccagatccgctgcagcagcaacatcacccgctgctgctgacccgc  
gacggcgccgaaggagatcagcaacaccaccgagatcttccgccccggcgccgacatgcgcgac  
aactggcgagcagctgtacaagtacaaggtgggtgaagatcgagccctgggctggccccacc  
atcgccatcagcagcgtggtgcagagcgagaagagcgcctgacccctgggcgccatgttctgggc  
ttctgggcgccgcccagcaccatgggcccgcagcctgacccctgacccgtgcaggccccgcccag  
ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgaggcccagcagcacctg  
ctgcagctgacccgtgtggggcatcaagcagctgcaggccccgctgctggcctggagcgtacctg  
aaggaccagcagctgctgggcatctggggctgcagcggcaagctgatctgcaccaccgcccgtgcc  
tggaacgccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtgg  
gagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggagagccagaaccagcag  
gagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtggaactggttcgacatc  
agcaagtggctgtggtacatctaactcgag

FIG. 33

(SEQ ID NO:46)

SUBSTITUTE SHEET (RULE 26)

45 / 131

gp140.mut8.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgcggtgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgcggtgaagctgacccccctgtgctgggcgccggcaactgccagacc  
agcgtgatcaccagggcctgcccccaagggtgagcttcgagcccatccccatccactactgcgcccc  
gcccgtctcgccatcctgaagtgcacaagaagttcaacggcagcggccccctgcaccaacgtg  
agcacctgtcagtgacacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagc  
ctggccgaggagggcggtgatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtg  
cagctgaaggagagcgtggagatcaactgcacccgcccccaacaacaacaccccgcaagagcatcacc  
atcgcccccgccgcgcttctacgccaccggcgacatcatcggcgacatccgccaggcccactgc  
aacatcagcggcgagaagtggacaacacccctgaagcagatcgtgaccaagctgcaggcccagttc  
ggcaacaagaccatcgtgttcaagcagagcagcggcgggcgacccccgagatcgtgatgcacagcttc  
aactgcggcgggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacacc  
atcgcccccaacaacaccaacggcaccatcaccttgccctgcccgcacatcaagcagatcatcaaccgc  
tggcaggaggtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaac  
atcacccgctgctgctgacccgcgacggcgggcaaggagatcagcaaacaccaccgagatcttccgc  
cccggcgggcgacatgcgcgacaactggcgcgagcagctgtacaagtacaagggtggtgaagatc  
gagccccctggcgctggccccaccatcgccatcagcagcgtggtgcagagcgagaagagcgccgtg  
accctgggcgcatgttccctgggcttccctgggcgcccggcgagcaccatgggcgccccgagcctg  
accctgaccgtgcaggccccgacgtgctgagcggcatcgtgcagcagcagaacaacctgctgcgc  
gccatcgaggcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggcccgc  
gtgctggccgtggagcgctacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag  
ctgatctgcaccaccgcccgtgcccgggaacgccagctggagcaacaagagcctggaccagatctgg  
aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctg  
atcgaggagagccagaaccagcaggagaagaacgagcagagctgctggagctggacaagtgggccc  
agcctgtggaactggttcgacatcagcaagtggctgtggtacatctaactcgag

**FIG. 34**

(SEQ ID NO:47)

SUBSTITUTE SHEET (RULE 26)

46 / 131

gp160.modSF162

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgaccctgcaactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgcggcgagatcaagaactgc  
agcttcaaggtagaccaccagcatccgcaacaagatgcagaaggagtacgccctgttctacaagctg  
gacgtggtgcccacgacaacgacaacaccagctacaagctgatcaactgcaacaccagcgtgatc  
acccaggcctgcccgaaggtagcttcgagcccatcccatccactactgccccccgcggcttc  
gcatcctggaagtgcacgacaagaagttaacggcagcgccccctgcaccaacgtgagcacccgtg  
cagtgaccccaacggcatccgccccgtggtgagcaccagctgctgctgaacggcagcctggccgag  
gagggcggtggtgatccgcagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaag  
gagagcgtggagatcaactgcacccgcccccaacaacaacaccgcgaagagcatcaccatcgcccc  
ggccgcgccttctacgccaccggcgacatcatcgggcgacatccgccaggcccaactgcaacatcagc  
ggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggcccaagtccggcaacaag  
accatcgtgttcaagcagagcagcgggcgaccccgagatcgtgatgcacagcttcaactgcggc  
ggcgagttcttctactgcaacagcaccagctgttcaacagcacctggaacaacaccatcgcccc  
aacaacaccaacggcaccatcacctgccccctgcccgcacatcaagcagatcatcaaccgctggcaggag  
gtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaacatcacggc  
ctgctgctgaccccgagcggcgccgaaggagatcagcaacaccaccgagatcttccgccccggcggc  
ggcgacatgcgcgacaactggcgagcagctgtacaagtacaagggtggaagatcgagccccctg  
ggcggtggcccccaacaggccaagcgccgctggtgagcgcgagaagcgccgtgaccctgggc  
gcatgttccctgggcttccctgggcgcccgcggcagcaccatgggcgcccgcagcctgaccctgacc  
gtgcaggccccgcagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcgccatcgag  
gcccagcagcacctgctgcagctgaccgtgtggggcatcaagcagctgcaggccccgcgtgctggcc  
gtggagcgctacctgaaggaccagcagctgctggggcatctggggctgcagcggcaagctgatctgc  
accaccgcccgtgcccctggaacgccagctggagcaacaagagcctggaccagatctggaacaacatg  
acctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctgatcgaggag  
agccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtgg  
aactggttcgacatcagcaagtggctgtggtacatcaagatcttcatcatgatcgtgggcggcctg  
gtgggcctgcgcacatcgtgttcaccgtgctgagcatcgtgaaccgcgtgcgccagggtacagcccc  
ctgagcttccagacccgcttccccgccccccgcggccccgacgccccgagggcatcgaggaggag  
ggcggcgagcgcgacgcgaccgcagcagccccctggtgcacggcctgctggccctgatctgggac  
gacctgcgcagcctgtgcctgttcagctaccaccgctgcgcgacctgatcctgatcgccgccccgc  
atcgtggagctgctggccgcccgcggctgggaggccctgaagtactggggcaacctgctgcagtac  
tggatccaggagctgaagaacagcgccgtgagcctgttcgacgccatcgccatcgccgtggccgag  
ggcaccgaccgcatcatcgaggtggcccagcgcacggccgcgccttctgcacatcccccgccgc  
atccgccagggttcgagcgcgcctgctgtaactcgag

**FIG. 35**

(SEQ ID NO:48)

SUBSTITUTE SHEET (RULE 26)



47 / 131

gp160.modSF162.delV2

gaattcgccaccatggatgcaatgaagagagggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgtgcccgtgtggaag  
gaggccaccaccacctgttctgcccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tgggcccaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctgacccctgactgcaccaacctg  
aagaacgccaccaacaccaagagcagcaactggaaggagatggaccgaggcgagatcaagaactgc  
agcttcaaggtgggcgccggcaagctgatcaactgcaacaccagcgtgatcaccaggcctgcccc  
aaggtgagcttcgagccccatccccatccactactgcgcccccgccggcttcgccatcctgaagtgc  
aacgacaagaagtccaacggcgagcgccccctgcaccaacgtgagcaccgtgcagtgacccacggc  
atccgccccgtggtgagcaccagctgctgctgaacggcagcctggccgaggagggcggtggtgatc  
cgcagcgagaacttcaccgacaacgccaagaccatcatcgtgcagctgaaggagagcgtggagatc  
gccaccggcgacatcatcgcgacatccgcccaggccccactgcaacatcagcggcgagaagtggaac  
aacaccctgaagcagatcgtgaccaagctgcaggccccagttcggaacaagaccatcgtgttcaag  
cagagcagcggcgagccccgagatcgtgatgcacagcttcaactgcggcgaggagttcttctac  
tgcaacagcaccagctgttcaacagcacctggaacaacacccatcgcccccaacaacaccaacggc  
accatcacccctgcccctgcccctcaagcagatcctcaaccgctggcaggaggtgggcaaggccatg  
tacgccccccccatccgcccagatccgctgcagcagcaacatcacccgcccgtgctgctgacccgc  
gacggcggaaggagatcagcaaacaccagagatctccgccccggcgggcgagatgcgagac  
aactggcgagcagcgtgtacaagtacaaggtggtgaagatcgagccccggggcggtggccccacc  
aaggccaagcgccgctggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc  
ttcctggggcgccgcccagcaccatggggcgcccgagcctgacccctgacccgtgcaggccccag  
ctgctgagcggcatcgtgcagcagcagaacaacctgctgcgcccacatcgaggccccagcagcagc  
ctgcagctgacccgtgtggggcatcaagcagctgcaggccccgctgctggcggtggagcgtacctg  
aaggaccagcagctgctggggcatctggggctgcagcggcaagctgatctgaccaccgcccgtgccc  
tggaacgccagctggagcaacaagagcctggaccagatctggaacaacatgacctggatggagtgg  
gagcgcgagatcgacaactacaccaacctgatctacaccctgatcgaggagagccagaaccagcag  
gagaagaacgagcaggagctgctggagctggacaagtgggcccagcctgtggaactggttcgacatc  
agcaagtggctgtggtacatcaagatcttcatcatgatcgtggggcgccctggtggggcctgcgcatc  
gtgttcaccgtgctgagcatcgtgaaccgctgcccagggctacagccccctgagcttccagacc  
cgcttccccgcccccgccgcccagcggccccgagggcatcgaggaggagggcgaggcgagcgcgac  
cgcgaccgcagcagccccctggtgcacggcctgctggccctgatctgggaagacctgcgcagcctg  
tgccgtgttcagctaccaccgctgcccagcctgatcctgatcgccgcccccatcgtggagctgctg  
ggcgccgaggctgggaggccctgaagtactggggcaacctgctgcagtaactggatccaggagctg  
aagaacagcggcgtgagcctgttcgacgccatcgccatcgccgtggcgaggcgaccgacgcacatc  
atcgaggtggccagcgcacggccgccccttctgcacatcccccgccgcatccgcccagggttc  
gagcgcgcccctgctgtaactcgag

FIG. 36

(SEQ ID NO:49)

SUBSTITUTE SHEET (RULE 26)

48 / 131

gp160.modSF162.delV1V2

gaattcgccaccatggatgcaatgaagagaggggctctgctgtgtgctgctgctgtgtggagcagtc  
ttcgtttcgcccagcgccgtggagaagctgtgggtgaccgtgtactacggcgctgcccgtgtggaag  
gagggcaccaccaccctgttctgcgccagcgacgccaaggcctacgacaccgaggtgcacaacgtg  
tggggcaccacgcctgctgcccaccgaccccaacccccaggagatcgtgctggagaacgtgacc  
gagaacttcaacatgtggaagaacaacatgggtggagcagatgcacgaggacatcatcagcctgtgg  
gaccagagcctgaagccctgctggaagctgacccccctgtgctggggcgccggcaactgccagacc  
agcgtgatcaccaggcctgccccagggtgagcttcgagccccatccccatccactactgcgcccc  
gcccgttcgcccacctgaagtgcacgacaagaagtcaacggcagcgccctgcaccaacgtg  
agcaccgtgcagtgacccacggcatccgccccgtggtgagcaccagctgctgctgaacggcagc  
ctggccgaggagggcgctggtgatccgcagcgagaacttcaccgacaacgcgaagaccatcatcgtg  
cagctgaaggagagcgtggagatcaactgcacccgcccccaacaacaacacccgcaagagcatcacc  
atcgcccccgccgcgcttctacgccaccggcgacatcatcgccgacatccgccaggcccaactgc  
aacatcagcggcgagaagtggaaacaacacctgaagcagatcgtgaccaagctgcaggcccaagtcc  
ggcaacaagaccatcgtgttcaagcagagcagcgccggcgacccccgagatcgtgatgcacagcttc  
aactgcgccggcgaggttcttctactgcaacagcaccagctgttcaacagcacctggaacaacacc  
atcgcccccaacaacaacagcgaccatcaccctgcccctgcccgcacagcagatcatcaaacgc  
tggcaggaggtgggcaaggccatgtacgccccccccatccgcggccagatccgctgcagcagcaac  
atcacgggctgctgctgacccgcgacggcgggcaaggagatcagcaacaccacagagatcttcgc  
ccccggcgccgagacatgcgcgacaactggcgcgagcagctgtacaagtacaagggtggtgaagatc  
gagccccctggcgctggccccccaccaaggccaaagcgccgcgtggtgcagcgcgagaagcgcgccgtg  
accctggcgcccatgttccctgggcttccctgggcgccgcccggcagcaccatggcgcccgagcctg  
accctgaccgtgcaggcccgccagctgctgagcggcatcgtgcagcagcagaacaacctgctgcgc  
gccatcgaggcccgagcagcaccctgctgcagctgaccgtgtggggcatcaagcagctgcaggcccg  
gtgctggccgtggagcgtacctgaaggaccagcagctgctgggcatctggggctgcagcggcaag  
ctgatctgcaccaccgcccgtgcccggaaacgcccagctggagcaacaagagcctggaccagatctgg  
aacaacatgacctggatggagtgggagcgcgagatcgacaactacaccaacctgatctacacctg  
atcgaggagagccagaaccagcaggagaagaacgagcaggagctgctggagctggacaagtgggoc  
agcctgtggaactggttcgacatcagcaagtggctgtgtgtacatcaagatcttcatcatgatcgtg  
ggcgccctggtgggcctgcgcacatcgtgttcaccgtgctgagcatcgtgaaccgcgtgcgccagggc  
tacagccccctgagcttccagaccgcttccccgcccccgccggccccgacgccccgagggcatc  
gaggaggagggcgccgagcgcgacccgcgaaccgcagcagccccctggtgcacggccctgctggccctg  
atctgggaacgacctgcgcagcctgtgacctgttcagctaccaccgcccgcgaacctgatcctgatc  
gcccggcgcatcgtggagctgctgggcccgcgcggctgggaggccctgaagtactggggcaacctg  
ctgcagtactggatccaggagctgaagaacagcgccgtgagcctgttcgacgcccacgcccacgccc  
gtggccgagggcaccgaccgcatcatcgaggtggcccagcgcatcgcccgccgcttctgcacatc  
ccccgcccgcacccgcccagggttcgagcgcgcacctgctgtaactcgag

**FIG. 37**  
(SEQ ID NO:50)

*SUBSTITUTE SHEET (RULE 26)*

49 / 131

## gp120wtUS4

ACAACAGTCTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG  
CAACCACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC  
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG  
GAAGTAAATTTAACAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACA  
TGGTGGAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAA  
GCCATGTGTAAAATTAACCCCACTCTGTGTTACTTTAAATTGTAAGTATAAGT  
TGACAGGTAGTACTAATGGCACAATAAGTACTAGTGGCACTAATAGTACTAG  
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA  
GAAGGAGAAATAAAAACTGCTCTTTCAATATCACCACAAGTGTAAGAGATA  
AAGTGCAGAAAGAATATTCTCTCTTCTATAAACTTGATGTAGTACCAATAGAT  
AATGATAATGCTAGCTATAGATTGATAAATTGTAATACCTCAGTCATTACACA  
AGCCTGTCCAAAGGTATCTTTTGAACCAATCCCATAATTATTGTGCCCCGG  
CTGGTTTTGCGATTCTAAAGTGTAAGATAAGAAGTTCAATGGAACAGGACC  
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA  
TCAACTCAACTGCTGTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA  
GATCTGAAAATTTACAGACAATGCTAAAACCATAATAGTACAGCTGAATGA  
ATCTGTAGAAATTAATTGTATAAGACCCAACAATAATACAAGAAAAAGTATA  
CATATAGGACCAGGGAGAGCATTATGCAACAGGTGATATAATAGGAGACA  
TAAGACAAGCACATTGTAACATTAGTAAAGCAAACCTGGACTAACACTTTAGA  
ACAGATAGTTGAAAAATTAAGAGAACAATTTGGGAATAATAAAACAATAATC  
TTTAATTCATCCTCAGGAGGGGACCCAGAAATTGTATTTACAGTTTTAATTG  
TGGAGGGGAATTTTTCTATTGTAATACATCAAACTATTTAATAGTACCTGGA  
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC  
ATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGGAAAAGCAAT  
GTATGCCCCCTCCCATCAGAGGACAAATTAATGTTTCATCAAATATTACAGGG  
CTGCTATTAAGTATAGATGGTGGTACTAACAATAATAGGACGAACGACACCG  
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT  
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCCA  
GGCAAAGAGAAGAGTGGTGCAAAGAGAGAAAAGA

## FIG. 38

(SEQ ID NO:51)

SUBSTITUTE SHEET (RULE 26)

50 / 131

gp140wtUS4

ACAACAGTCTTGTGGGTACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG  
CAACCACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC  
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG  
GAAGTAAATTTAACAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACA  
TGGTGGAAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAA  
GCCATGTGTAAAATTAACCCCACTCTGTGTTACTTTAAATTGTAAGTACTGATAAGT  
TGACAGGTAGTACTAATGGCACAATAAGTACTAGTGGCACTAATAGTACTAG  
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA  
GAAGGAGAAAATAAAAACTGCTCTTTCAATATCACCACAAGTGTAAGAGATA  
AAGTGCAGAAAGAATATTCTCTCTTCTATAAACTTGATGTAGTACCAATAGAT  
AATGATAATGCTAGCTATAGATTGATAAATTGTAATACTCAGTCATTACACA  
AGCCTGTCCAAAGGTATCTTTGAACCAATTCCCATACATTATTGTGCCCCGG  
CTGGTTTTGCGATTCTAAAGTGTAAGATAAGAAGTTCAATGGAACAGGACC  
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA  
TCAACTCAACTGCTGTTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA  
GATCTGAAAATTTACAGACAATGCTAAAACCATAATAGTACAGCTGAATGA  
ATCTGTAGAAATTAATTGTATAAGACCCAACAATAATACAAGAAAAAGTATA  
CATATAGGACCAGGGAGAGCATTTTATGCAACAGGTGATATAATAGGAGACA  
TAAGACAAGCACATTGTAACATTAGTAAAGCAAACCTGGACTAACACTTTAGA  
ACAGATAGTTGAAAAATTAAGAGAACAAATTTGGGAATAATAAAACAATAATC  
TTTAATTCATCCTCAGGAGGGGACCCAGAAATTGTATTTACAGTTTTAATTG  
TGGAGGGGAATTTTCTATTGTAATACATCACAACCTATTTAATAGTACCTGGA  
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC  
ATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGGAAAAGCAAT  
GTATGCCCCTCCCATCAGAGGACAAATTAAATGTTTCATCAAATATTACAGGG  
CTGCTATTAAGTATAGAGATGGTGGTACTAACAATAATAGGACGAACGACACCG  
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT  
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCCA  
GGCAAAGAGAAGAGTGGTGCAAAGAGAGAAAAGAGCAGTGGGACTAGGAG  
CTTTGTTTCATTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTC  
AGTGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAACAG  
CAGAACAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCA  
CGGTCTGGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATA  
CCTAAAGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACCTATTT  
GCACCACTACTGTGCCTTGGAACCTAGTTGGAGTAATAAATCTCTGACTGAG  
ATTTGGGATAATATGACCTGGATGGAGTGGGAAAGAGAAAATTGGCAATTATA  
CAGGCTTAATATACAATTTAATTGAAATAGCACAAAACCAGCAAGAAAAGAA  
TGAACAAGAATTATTGGAATTAGACAAGTGGGCAAGTTTGTGGAATTGGTTT  
GATATAACAACTGGCTGTGGTATATA

**FIG. 39**

(SEQ ID NO:52)

SUBSTITUTE SHEET (RULE 26)

51 / 131

gp160wtUS4

ACAACAGTCTTGTGGGTCACAGTCTATTATGGGGTACCTGTGTGGAAAGAAG  
CAACCACCACTCTGTTTTGTGCATCAGATGCTAAAGCATACAAAGCAGAGGC  
ACATAACGTCTGGGCTACACATGCCTGTGTACCCACAGACCCCAACCCACAG  
GAAGTAAATTTAACAAATGTGACAGAAAATTTTAACATGTGGAAAAATAACA  
TGGTGGAAACAGATGCATGAGGATATAATCAGTTTATGGGATCAAAGCCTAAA  
GCCATGTGTAAAATTAACCCCACTCTGTGTTACTTTAAATTGTAAGTACTAG  
TGACAGGTAGTACTAATGGCACAAATAGTACTAGTGGCACTAATAGTACTAG  
TGGCACTAATAGTACTAGTACTAATAGTACTGATAGTTGGGAAAAGATGCCA  
GAAGGAGAAATAAAAACTGCTCTTTCAATATCACCACAAGTGTAAGAGATA  
AAGTGCAGAAAGAATATTCTCTCTTCTATAAACTTGATGTAGTACCAATAGAT  
AATGATAATGCTAGCTATAGATTGATAAATTGTAATACCTCAGTCATTACACA  
AGCCTGTCCAAAGGTATCTTTTGAACCAATTCCCATAACATTATTGTGCCCCGG  
CTGGTTTTGCGATTCTAAAGTGTAAGATAAGAAGTTCAATGGAACAGGACC  
ATGTAAAAATGTCAGCACAGTACAATGCACACATGGAATTAGACCAGTAGTA  
TCAACTCACTGCTGTAAATGGCAGTCTAGCAGAAGAAGAGATAGTACTTA  
GATCTGAAAATTTACAGACAATGCTAAAACCAATAATAGTACAGCTGAATGA  
ATCTGTAGAAATTAATTGTATAAGACCCAAACAATAACAAGAAAAAGTATA  
CATATAGGACCAGGGAGAGCATTTTATGCAACAGGTGATATAATAGGAGACA  
TAAGACAAGCACATTGTAACATTAGTAAAGCAAACCTGGACTAACACTTTAGA  
ACAGATAGTTGAAAAATTAAGAGAACAAATTTGGGAATAATAAAACAATAATC  
TTTAATTCATCCTCAGGAGGGGACCCAGAAATTGTATTTACAGTTTTAATTG  
TGGAGGGGAATTTTTCTATTGTAATACATCACAATTTAATAGTACCTGGA  
ATATTACTGAAGAGGTAAATAAGACTAAAGAAAATGACACTATCATACTCCC  
ATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGGAAAAGCAAT  
GTATGCCCTCCCATCAGAGGACAAATTAATGTTTCATCAAATATTACAGGG  
CTGCTATTAAGTAGAGATGGTGGTACTAACAAATAATAGGACGAACGACCCG  
AGACCTTCAGACCTGGGGGAGGAAACATGAAGGACAATTGGAGAAGTGAAT  
TATATAAATATAAAGTAGTAAGAATTGAACCATTAGGAGTAGCACCCACCCA  
GGCAAAGAGAAGAGTGGTGCAAAGAGAGAAAAGAGCAGTGGGACTAGGAG  
CTTTGTTTATTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTC  
AGTGACGCTGACGGTACAGGCCAGACAATTATTGTCTGGTATAGTGCAACAG  
CAGAACAATTTGCTGAGAGCTATTGAGGCGCAACAGCATCTGTTGCAACTCA  
CGGTCTGGGGCATCAAACAGCTCCAGGCAAGAATCCTGGCTGTGGAAAGATA  
CCTAAAGGATCAACAGCTCCTAGGGATTTGGGGTTGCTCTGGAAAACCTATTT  
GCACCACTACTGTGCCTTGGAACTCTAGTTGGAGTAATAAATCTCTGACTGAG  
ATTTGGGATAATATGACCTGGATGGAGTGGGAAAGAGAAATTGGCAATTATA  
CAGGCTTAATATACAATTTAATTGAAATAGCACAAAACCAGCAAGAAAAGAA  
TGAACAAGAATTATTGGAATTAGACAAGTGGGCAAGTTTGTGGAATTGGTTT  
GATATAACAACTGGCTGTGGTATATAAGAATATTCATAATGATAGTAGGAG  
GCTTGATAGGTTTAAGAATAGTTTTTGTGCTGTTCTATAGTGAATAGAGTT  
AGGCAGGGATACTACCAATATCATTGCAGACCCGCTCCAGCTCAGAGGG

**FIG. 40A**

(SEQ ID NO:53)

SUBSTITUTE SHEET (RULE 26)

52 / 131

GACCCGACAGGCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGA  
GACAGATCCAATCGATTAGTGCATGGATTATTGGCACTCATCTGGGACGATCT  
GCGGAGCCTGTGCCTCTTCAGCTACCACCGCTTGAGAGACTTACTCTTGATTG  
TAGCGAGGATTGTGGAACCTTCTGGGACGCAGGGGGTGGGAAGCCCTCAAGTA  
TTGGTGGAATCTCCTGCAGTATTGGAGTCAGGAGCTAAAGAGTAGTGCTGTT  
AGTTTGTTTAATGCCACAGCAATAGCAGTAGCTGAAGGGACAGATAGGATTA  
TAGAAATAGTACAAAGAATTTTATAGAGCTGTAATTCACATACCTAGAAGAAT  
AAGACAGGGCTTGGAGAGGGCTTTACTATAA

**FIG. 40B**  
(SEQ ID NO:53)

*SUBSTITUTE SHEET (RULE 26)*

53 / 131

**gp120.modUS4**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTGCCCGTG  
TGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCGAGCGGCAC  
CAACAGCACCGCGGCACCAACAGCACCGACCAACAGCACCGACAGCTGGGAGAAGATG  
CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA  
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAACGCCAGCT  
ACCGCCTGATCAACTGCAACACCAGCGTGATCACCAGGCCTGCCCAAGGTGAGCTTCGAGC  
CCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGT  
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAAGTGCACCCACGGCATCCGCCCC  
GTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC  
CGAGAACTTCACCGACAACGCCAAGACCATCATCGTGCAAGTGAACGAGTCCGTGGAGATCA  
ACTGCATCCGCCCCAACAAACAACACGCGTAAGAGCATCCACATCGGCCCGGCCGCGCCTTCT  
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAC  
TGGACCAACACCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTTCGGCAACAACAAGAC  
CATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGG  
CGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGA  
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA  
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGC  
AGCAGCAATATTACCGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCA  
CGACACCGAGACCTTCCGCCCCGGCGGCGGAACATGAAGGACAACCTGGCGCAGCGAGCTGT  
ACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGCCCCCACCAGGCCAAGCGCCGC  
GTGGTGCAAGCGGAGAAGCGCTAAGATATCGGATCCTCTAGA

**FIG. 41**  
(SEQ ID NO:54)

*SUBSTITUTE SHEET (RULE 26)*

54 / 131

**gp120.mod.US4.del128-194**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG  
AGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCG  
TGCCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCAGCGACGCCAAGGCTTAC  
AAGGCCGAGGCCACAACTGTGGGCCACCCACGCCTGCGTGCCCAACGACCCCAACCC  
CCAGGAGGTGAACCTGACCAACGTGACCGAGAAGTTCAACATGTGGAAGAACAACATGG  
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG  
AAGCTGACCCCCCTGTGCGTGGGGGCAGGGAAGTGCAGAGACCAGCGTGATCACCAGGC  
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG  
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC  
ACCGTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGG  
CAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAAGTTACCGACAACGCCAAGA  
CCATCATCGTGCAAGTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAAC  
ACGCGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACGCCACCGGCGACATCAT  
CGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTCG  
AGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGCAACAACAAGACCATCATCTTCAAC  
AGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTT  
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCAGGAGGTGA  
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAAC  
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTG  
CAGCAGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCA  
CCAACGACACCGAGACCTTCCGCCCCGGCGGCGGCAACATGAAGGACAAGTGGCGCAGC  
GAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGGCCCCACCCAGGC  
CAAGCGCCGCGTGGTGAGCGCGAGAAGCGCTAAGATATCGGATCCTCTAGA

**FIG. 42**

(SEQ ID NO:55)

*SUBSTITUTE SHEET (RULE 26)*



55 / 131

gp140.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGACTACGGCGTGCCCGTG  
TGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGAGAAGCTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCTGAACTGCACCGACAAGCTGACCGGCAGCACCACCGCACCAACAGCACCAGCGGCAC  
CAACAGCACCAGCGGCACCAACAGCACCAGCACCACAGCACCAGCCTGCGCGACAAGGTGCA  
CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA  
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAACGCCAGCT  
ACCGCCTGATCAACTGCAACACCAGCGTGATCACCAGGCCTGCCCAAGGTGAGCTTCGAG  
CCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGT  
TCAACGGCACCGGGCCCTGCAAGAAGCTGAGCACCCTGCAAGTGACCCACGGCATCCGCCCC  
GTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC  
CGAGAACTTCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGAGTCCGTGGAGATCA  
ACTGCATCCGCCCCAACAACAACACGCGTAAGAGCATCCACATCGGCCCGGCCGCGCTTCT  
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC  
TGGACCAACACCCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGCAACAACAAGAC  
CATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGG  
CGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCAGGA  
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCA TCCGCCAGATCATCA  
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGC  
AGCAGCAATATTACCGGCCTGTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCA  
CGACACCGAGACCTTCCGCCCCGGCGGCGCAACATGAAGGACAACCTGGCGCAGCGAGCTGT  
ACAAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTGCCCCCAACCAAGCGCCGC  
GTGGTGACGCGGAGAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTCATCGGCTTCTGGGCGC  
GCCGGGAGCACCATGGGCGCCGCTCCGTGACCCGTGACCGTGACGGCCCGCCAGCTGCTGAG  
CGGCATCGTGACGAGCAGACAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCCTGCTGC  
AGCTGACCGTGTTGGGCGCATCAAGCAGCTGACGGCCCGCATCCTGGCCGTGGAGCGCTACCTG  
AAGGACCAGCAGCTGCTGGGCATCTGGGCTGACGGCGCAAGCTGATCTGCACCACCACCGT  
GCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA  
TGGAGTGGGAGCGGAGATCGGCAACTACCCGGCCTGATCTACAACCTGATCGAGATCGCC  
CAGAACCAGCAGGAGAAGAAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT  
GGAAGTGGTTCGACATCAACCACTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

FIG. 43

(SEQ ID NO:56)

SUBSTITUTE SHEET (RULE 26)

56 / 131

**gp140.mut.modUS4**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCAACCGTGCTGTGGGTGACCGTGTAACGCGGTGCCCGTG  
TGGAAAGGAGGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGAGAATTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCGCGGCAC  
CAACAGCACCGCGCACCAAGCAGCACCAACAGCACCGACAGCTGGGAGAAGATG  
CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCAACAGCGTGCGCGACAAGGTGCA  
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAACGCCAGCT  
ACCGCCTGATCAACTGCAACACCAGCGTGATCACCCAGGCCTGCCCAAGGTGAGCTTCGAGC  
CCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGT  
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAAGTGACCCACGGCATCCGCCCC  
GTGGTGAGCAGCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC  
CGAGAATTACCGACAACGCCAAGACCATCATCGTGCAAGTGAACGAGTCCGTGGAGATCA  
ACTGCATCCGCCCCAACAACAACGCGTAAGAGCATCCACATCGGCCCGCGCGCCTTCT  
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAC  
TGGACCAACACCCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGAC  
CATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGG  
CGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGA  
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCA TCCGCCAGATCATCA  
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGC  
AGCAGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCA  
CGACACCGAGACCTTCCGCCCCGGCGGCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTGT  
ACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGGCCCCACCCAGGCCAAGCGCCGC  
GTGGTGAGCGCGAGAAGAGCGCCGTGGGCGCCTGTTTCATCGGCTTCCTGGGCGCC  
GCCGGGAGCACCATGGGCGCCGCTCCGTGACCTGACCGTGACGGCCCGCCAGCTGCTGAG  
CGGCATCGTGAGCAGCAGAACAACTGCTGCGCGCCATCGAGGCCAGCAGCACTGCTGC  
AGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCGCTACCTG  
AAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCAACACCGT  
GCCCTGGAAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA  
TGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTACAACCTGATCGAGATCGCC  
CAGAACAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT  
GGAAGTGGTTCGACATCACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTAGA

**FIG. 44**

(SEQ ID NO:57)

*SUBSTITUTE SHEET (RULE 26)*

57 / 131

**gp140.TM.modUS4**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACCGGCGTCCCGTG  
TGGAAGGAGGCCACCACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCCAGCGGCAC  
CAACAGCACCCAGCGGCACCAACAGCACCCAGCACCAACAGCACCCGACAGCTGGGAGAAGATG  
CCCGAGGGCGAGATCAAGAACTGCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA  
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAACGCCAGCT  
ACCGCCTGATCAACTGCAACACCAGCGTGATCACCAGGCCTGCCCAAGGTGAGCTTCGAGC  
CCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGT  
TCAACGGCACCCGGCCCCCTGCAAGAACGTGAGCACCGTGCAAGTGCAACCGCATCCGCCCC  
GTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC  
CGAGAATTCACCGACAACGCCAAGACCATCATCGTGACGTGAACGAGTCCGTGGAGATCA  
ACTGCATCCGCCCCAACAACAACACGCGTAAGAGCATCCACATCGGCCCGCGCCGCGCCTTCT  
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAAC  
TGGACCAACACCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGAC  
CATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGG  
CGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGA  
GGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCA  
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGC  
AGCAGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCA  
CGACACCGAGACCTTCCGCCCCGGCGGCGGAACATGAAGGACAACCTGGCGCAGCGAGCTGT  
ACAAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTGCCCCCACCCAGGCCAAGCGCCGC  
GTGGTGACGCGGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGTTTCATCGGCTTCTGGGCGCC  
GCCGGGAGCACCATGGGCGCCGCTCCGTGACCCTGACCGTGACGGCCCGCCAGCTGCTGAG  
CGGCATCGTGACGAGCAGAGAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCTGCTGC  
AGCTGACCGTGTTGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCGCTACCTG  
AAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGAAGCTGATCTGCACCACCACCGT  
GCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA  
TGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTACAACCTGATCGAGATCGCC  
CAGAACCAGCAGGAGAAGAAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT  
GGAACCTGGTTCGACATACCAACTGGCTGTGGTACATCCGCATCTTCATCATGATCGTGGGCG  
GCCTGATCGGCCTGCGCATCGTGTTCCGCGTGCTGAGCATCGTGTAAGATATCGGATCCTCTA  
GA

**FIG. 45**  
(SEQ ID NO:58)

*SUBSTITUTE SHEET (RULE 26)*

58 / 131

## Gp140modUS4.DV1V2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGC  
TGTGTGGAGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACC  
GTGTAACGCGTGTGCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCG  
CCAGCGACGCCAAGGCTTACAAGGCCAGGCCACAACGTGTGGGCCACCCA  
CGCCTGCGTGCCACCGACCCCAACCCCAAGGAGGTGAACCTGACCAACGTG  
ACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCGCCGGCC  
AGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCC  
CGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGC  
CCCTGCAAGAACGTGAGCACCGTGAGTGACCCACGGCATCCGCCCCGTGG  
TGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCT  
GCGCTCCGAGAACTTACCCGACAACGCCAAGACCATCATCGTGACGTGAAC  
GAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACGCGTAAGAGCA  
TCCACATCGGCCCCCGGCCGCGCCTTCTACGCCACCGGCGACATCATCGGCGA  
CATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTC  
GAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCCGGCAACAACAAGACCATC  
ATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCA  
ACTGCGGCGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCAC  
CTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT  
CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAAG  
GCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGACAGCAATATTA  
CCGGCCTGTGTGACCCGCGACGGCGGCAACCAACAACCGCACCAACGA  
CACCGAGACCTTCCGCCCCGGCGGCGGCAACATGAAGGACAACCTGGCGCAGC  
GAGCTGTACAAGTACAAGGTGGTGCATCGAGCCCCTGGGCGTGGCCCCCA  
CCCAGGCCAAGCGCCGCGTGGTGCAGCGGAGAAGCGCGCCGTGGGCCTGG  
GCGCCCTGTTTCATCGGCTTCCTGGGCGCCGCGGGAGCACCATGGGCGCCGC  
CTCCGTGACCCTGACCGTGACGGCCCGCCAGCTGCTGAGCGGCATCGTGACG  
CAGCAGAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCTGCTGCAGC  
TGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCG  
CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTG  
ATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGA  
CCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCA  
ACTACACCGGCCTGATCTACAACCTGATCGAGATCGCCAGAACCAGCAGGA  
GAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAA  
CTGGTTCGACATCACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTCTA  
GA

## FIG. 46

(SEQ ID NO:59)

SUBSTITUTE SHEET (RULE 26)

59 / 131

Gp140modUS4.DV2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGC  
TGTGTGGAGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACC  
GTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCG  
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCCA  
CGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGTG  
ACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCC  
CCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGG  
CACCAACAGCACCAGCGGCACCAACAGCACCAGCGGCACCAACAGCACCAG  
CACCAACAGCACCAGCAGCTGGGAGAAGATGCCCCAGGGCGAGATCAAGAA  
CTGCAGCTTCAACATCGGCGCCGGCCGCTGATCAACTGCAACACCAGCGTG  
ATCACCCAGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACT  
GCGCCCCCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGG  
CACCGGCCCTGCAAGAACGTGAGCACCCTGCAAGTGCAGTGCACCCACGGCATCCGC  
CCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGA  
TCGTGCTGCGTCCGAGAACTTCAACGACAACGCCAAGACCATCATCGTGCA  
GCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACGCGT  
AAGAGCATCCACATCGGCCCGCCGCGCCTTCTACGCCACCGGCGACATCA  
TCGGCGACATCCGCCAGGCCCCACTGCAACATCAGCAAGGCCAACTGGACCAA  
CACCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAA  
GACCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCAC  
AGCTTCAACTGCGGCGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAA  
CAGCACCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACAC  
CATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTG  
GGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCA  
ATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCAC  
CAACGACACCGAGACCTTCCGCCCGGCGGCGGCAACATGAAGGACAACCTG  
GCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTG  
GCCCCACCCAGGCCAAGCGCCGCGTGGTGACGCGGAGAAGCGCGCCGTG  
GGCCTGGGCGCCCTGTTTCATCGGCTTCCTGGGCGCCGCGGAGCACCATGG  
GCGCCGCTCCGTGACCCTGACCGTGACGGCCCGCCAGCTGCTGAGCGGCAT  
CGTGACAGCAGCAGAACAACTGCTGCGCGCCATCGAGGCCAGCAGCACCTG  
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCCGATCCTGGCCG  
TGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGG  
CAAGCTGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAG  
AGCCTGACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAG  
ATCGGCAACTACACCGCCTGATCTACAACCTGATCGAGATCGCCAGAACC  
AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCC  
TGTGGAACGTGTTTCGACATCACCAACTGGCTGTGGTACATCTAAGATATCGG  
ATCCTCTAGA

**FIG. 47**

(SEQ ID NO:60)

SUBSTITUTE SHEET (RULE 26)

60 / 131

**Gp140modmutUS4.DV1V2**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGC  
TGTGTGGAGCAGTCTTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACC  
GTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCG  
CCAGCGACGCCAAGGCTTACAAGGCCGAGGCCACAACGTGTGGGCCACCC  
ACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACCTGACCAACGT  
GACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGA  
GGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCGCCGGC  
CAGGCCTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCC  
CCGCCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCCG  
CCCCTGCAAGAACGTGAGCACCGTGCACTGCACCCACGGCATCCGCCCCGTG  
GTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGC  
TGCGCTCCGAGAACTTCACCGACAACGCCAAGACCATCATCGTGCAGCTGAA  
CGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACGCGTAAGAGC  
ATCCACATCGGCCCCGGCCGCGCCTTCTACGCCACCGGCGACATCATCGGCG  
ACATCCGCCAGGCCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCCT  
CGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCAT  
CATCTTCAACAGCAGCAGCGCGGCGACCCCGAGATCGTGTTCACAGCTTC  
AACTGCGGCGGCGAGTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCA  
CCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCA  
TCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAA  
GGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAATATT  
ACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCAACG  
ACACCGAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAACCTGGCGCA  
GCGAGCTGTACAAGTACAAGGTGGTGCAGCATCGAGCCCCTGGGCGTGGCCCC  
CACCCAGGCCAAGCGCCGCGTGGTGCAGCGGAGAGCGCCGTGGGCCT  
GGGCGCCCTGTTTCATCGGCTTCCTGGGCGCCGCGGGAGCACCATGGGCGCC  
GCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCTGAGCGGCATCGTGC  
AGCAGCAGAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCTGCTGCA  
GCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAG  
CGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGC  
TGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT  
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGG  
CAACTACACCGGCCTGATCTACAACCTGATCGAGATCGCCAGAACCAGCAG  
GAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGG  
AACTGGTTCGACATACCAACTGGCTGTGGTACATCTAAGATATCGGATCCTC  
TAGA

**FIG. 48**

(SEQ ID NO:61)

*SUBSTITUTE SHEET (RULE 26)*

51 / 131

gp140.mod.US4.del128-194

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTG  
AGCAGTCTTCGTTTCGCCACGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCG  
TGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTAC  
AAGGCCGAGGCCCAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCC  
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGG  
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG  
AAGCTGACCCCCCTGTGCGTGGGGGCAGGGAAGTGCAGAGACCAGCGTGATCACCAGGC  
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCG  
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAAGTGAGC  
ACCGTGCAAGTGACCCACGGCATCCGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGG  
CAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAAGTTCACCGACAACGCCAAGA  
CCATCATCGTGAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAAC  
ACGCGTAAGAGCATCCACATCGGCCCCCGCGCCTTCTACGCCACCGGCGACATCAT  
CGGCGACATCCGCCAGGCCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTCG  
AGCAGATCGTGGAAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC  
AGCAGCAGCGGCGGCGACCCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTT  
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCAGGAGGTGA  
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAAC  
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTG  
CAGCAGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCA  
CCAACGACACCGAGACCTTCCGCCCCGCGGCGGCAACATGAAGGACAAGTGGCGCAGC  
GAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGCCCCCACCAGGC  
CAAGCGCCGCGTGGTGAGCGCGAGAAGCGCGCCGTGGGCGTGGCGCCCTGTTTCATCG  
GCTTCCTGGGCGCCGCGGAGCACCATGGGCGCCGCTCCGTGACCCGTGACCGTGACG  
GCCCCAGCAGCACCTGCTGAGCGGCATCGTGAGCAGCAGAACAACCTGCTGCGCGCCATCGA  
GGCCCAGCAGCACCTGCTGAGCGCTGACCGTGTTGGGCGATCAAGCAGCTGCAGGCCCGCA  
TCCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCGATCTGGGGCTGCAGC  
GGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT  
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACA  
CCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACCAGCAGGAGAAGAAGAGCAG  
GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAAGTGGTTTCGACATCACCAACTG  
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

**FIG. 49**

(SEQ ID NO:62)

SUBSTITUTE SHEET (RULE 26)

62/131

gp140.mut.mod.US4.del128-194

GAATTGCCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGG  
AGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTACTACGGCG  
TGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTAC  
AAGGCCGAGGCCCAACCGTGTGGGCCACCCACGCCTGCGTGCCCAACGACCCCAACCC  
CCAGGAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGG  
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG  
AAGCTGACCCCCCTGTGCGTGGGGGCAGGGAACTGCGAGACCAGCGTGATCACCCAGGC  
CTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCG  
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGC  
ACCGTGCAAGTGACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGG  
CAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTCACCGACAACGCCAAGA  
CCATCATCGTGCAAGTCCCGTGGAGATCAACTGCATCCGCCCCAACAACAAC  
ACGCGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACGCCACCGGCGACATCAT  
CGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCCTCG  
AGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAAC  
AGCAGCAGCGGCGGCGACCCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTT  
CTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGA  
ACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCAAC  
ATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTG  
CAGCAGCAATATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCA  
CCAACGACACCGAGACCTTCCGCCCCGGCGGCGGCAACATGAAGGACAACCTGGCGCAGC  
GAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTGCCCCCACCAGGC  
CAAGCGCCGCGTGGTGACGCGGAGAGAGCGCCGTGGGCCTGGGCGCCCTGTTTCATCG  
GCTTCCTGGGCGCCGCGGGAGCACCATGGGCGCCGCTCCGTGACCCTGACCGTGACG  
GCCCCGAGCTGCTGAGCGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGA  
GGCCCAGCAGCACCTGCTGACGCTGACCGTGTGGGCATCAAGCAGCTGCAGGCCCGCA  
TCCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGC  
GGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCT  
GACCGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACA  
CCGGCCTGATCTACAACCTGATCGAGATCGCCAGAACCAGCAGGAGAAGAACGAGCAG  
GAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTG  
GCTGTGGTACATCTAAGATATCGGATCCTCTAGA

**FIG. 50**

(SEQ ID NO:63)

*SUBSTITUTE SHEET (RULE 26)*



63/131

## gp160.modUS4

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTGCCCCGTG  
TGGAAGGAGGGCCACCACCCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGAGAACTTCAACATGTGGAAGAAACAATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACCAACAGCACCAGCGGCAC  
CAACAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACCAGCTGGGAGAAGATG  
CCCGAGGGCGAGATCAAGAAGCTGACGTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA  
GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACGACAACGCCAGCT  
ACCGCCTGATCAACTGCAACACCAGCGTGATCACCCAGGCCTGCCCCAAGGTGAGCTTCGAGC  
CCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGT  
TCAACGGCACCGGCCCTGCAAGAACGTGAGCACCCTGCAAGTGCAACCGGCATCCGCCCC  
GTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTC  
CGAGAACTTCACCGACAACGCCAAGACCATCATCGTGACGCTGAACGAGTCCGTGGAGATCA  
ACTGCATCCGCCCCAACAACAACCGCTAAGAGCATCCACATCGGCCCGGCCGCGCTTCT  
ACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAAC  
TGGACCAACACCCCTCGAGCAGATCGTGGAGAAAGCTGCGCGAGCAGTTGCGCAACAACAAGAC  
CATCATCTTCAACAGCAGCAGCGCGCGCAACCCGAGATCGTGTTCCACAGCTTCAACTGCGG  
CGGCGAGTTCTTACTGCAACACCAGCCAGCTGTTCAACAGCAGCTGGAACATCACCGAGGA  
GGTGAACAAGACCAAGGAGAACGACACCATCATCTGCCCTGCCGATCCGCCAGATCATCA  
ACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGC  
AGCAGCAATATTACCGCCTGCTGCTGACCCGCGACGCGGCCACCAACAACAACCGCACCA  
CGACACCGAGACCTTCCGCCCCGCGCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTGT  
ACAAGTACAAGGTGGTGGCATCGAGCCCTGGGCGTGGCCCCCAACAGGCCAAGCGCCGC  
GTGGTGACGCGGAGAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTATCGGCTTCTGGGCGCC  
GCCGGGAGCACCATGGGCGCCGCTCCGTGACCCCTGACCGTGACGGCCCGCCAGCTGCTGAG  
CGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGAGGCCCGCCAGCTGCTGAG  
AGCTGACCGTGTGGGCGATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCGCTACCTG  
AAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCACCGT  
GCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGA  
TGGAGTGGGAGCGCGAGATCGGCAACTACCCGGCCTGATCTACAACCTGATCGAGATCGCC  
CAGAACCAGCAGGAGAAGAAGCAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGT  
GGAACCTGGTTCGACATCACCAACTGGCTGTGGTACATCCGCATCTTATCATGATCGTGGGCG  
GCCTGATCGGCCTGCGCATCGTGTTGCGCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCT  
ACAGCCCCATCAGCCTGCAGACCCGCTGCCCGCCAGCGCGGCCCGACCGCCCCGAGGGC  
ATCGAGGAGGAGGGCGGCGAGCGGACCGCGACCGCAGCAACCGCCTGGTGCACGGCCTGCT  
GGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCTGTTAGCTACCAACCGCCTGCGCGACCT  
GCTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCCGCGCGGCTGGGAGGCCCTGAAGT  
ACTGGTGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTT  
AACGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCAT  
CTTCCGCGCCGTGATCCACATCCCCCGCGCATCCGCCAGGGCCTGGA GCGCGCCCTGCTGTA  
AGATATCGGATCCTCTAGA

FIG. 51

(SEQ ID NO:64)

SUBSTITUTE SHEET (RULE 26)

64 / 131

**gp160.modUS4.delV1**

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG  
TGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCAAGGAGGTGAACC  
TGACCAACGTGACCGAGAACTTCAACATGTGGGAAGAACACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
ACCCTGAACTGCACCGACAAGCTGGGCGCCGGCGGCGAGATCAAGAACTGCAGCTTCAACAT  
CACCACCAGCGTGCGCGACAAGGTGCAGAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGG  
TGCCCATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACCAGCGTGATCACCC  
AGGCCTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCG  
CCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGCACC  
GTGCAGTGCACCCACGGCATCCGCCCCGTGGTGAGCAACCCAGCTGCTGCTGAACGGCAGCCTG  
GCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTACCGACAACGCCAAGACCATCATCGT  
GCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAACACGCGTAAGAGCA  
TCCACATCGCCCCCGCGCGCCTTCTACGCCACCGGCGACATCATCGGCGACATCCGCCAGG  
CCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAGATCGTGGAGAAGCTG  
CGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAACAGCAGCAGCGCGCGGACCCCGA  
GATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTT  
CAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCC  
TGCCCTGCCGCATCCGCCAGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGACCCGCGAC  
CCCCCATCCCGCGCCAGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGACCCGCGAC  
GGCGGCACCAACAACAACCGCACCAACGACACCGAGACCTTCCGCCCCGGCGGCGGCAACAT  
GAAGGACAACCTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGGCGATCGAGCCCCCTGGGCG  
TGGCCCCCACCAGGCCAAGCGCCGCGTGGTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGC  
GCCCTGTTTCATCGGCTTCTGGGCGCCGCGGAGCACCATGGGCGCCGCTCCGTGACCCCTG  
ACCGTGCAGGCCCCGAGCTGCTGAGCGGCATCGTGACGAGCAGAGCAACAACCTGCTGCGCGC  
CATCGAGGCCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCC  
GCATCCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGC  
GGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGAC  
CGAGATCTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCC  
TGATCTACAACCTGATCGAGATCGCCCAGAACCAGCAGGAGAAGAACGAGCAGGAGCTGCTG  
GAGCTGGACAAGTGGGCCAGCCTGTGGAACTGGTTTCGACATCACCAACTGGCTGTGGTACATC  
CGCATCTTCATCATGATCGTGGGCGGCTGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGC  
ATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCCTGCAGACCCGCTGCCCGCCCAG  
CGCGGCCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCGCGACCGCGACCGCA  
GCAACCGCCTGGTGACGGCCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCCTGT  
TCAGCTACCACCGCCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCC  
GCCGCGGCTGGGAGGCCCTGAAGTACTGGTGGAACTGCTGCAGTACTGGAGCCAGGAGCTG  
AAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCG  
CATCATCGAGATCGTGCAGCGCATCTTCCGCGCCGTGATCCACATCCCCCGCGCATCCGCCA  
GGCCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA

**FIG. 52**

(SEQ ID NO:65)

*SUBSTITUTE SHEET (RULE 26)*

65 / 131

## gp160.mod.US4.delV2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG  
AGCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGACTACGGCG  
TGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTAC  
AAGGCCGAGGCCCAACCGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCC  
CCAGGAGGTGAACCTGACCAACGTGACCGAGAATTCAACATGTGGAAGAACAACATGG  
TGGAGCAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG  
AAGCTGACCCCCCTGTGCGTGACCTGAACTGCACCGACAAGCTGACCGGCAGCACCAA  
CGGCACCAACAGCACCAGCGGCACCAACAGCACCAGCGGCACCAACAGCACCAGCACCA  
ACAGCACCAGCAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACTGCAGCTTCAAC  
ATCGGCGCGCGGCCCTGATCAACTGCAACACCAGCGTGATCACCAGGCCTGCCCCAA  
GGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGA  
AGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAACGTGAGCACCCTGCG  
TGCACCCACGGCATCCGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGC  
CGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTCACCGACAACGCCAAGACCATCATCG  
TGCAGCTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACGCGTAAG  
AGCATCCACATCGGCCCGCGCGCCTTCTACGCCACCGGCGACATCATCGGCGACAT  
CCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAGATCG  
TGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAACAGCAGCAGC  
GGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTTCTTCTACTG  
CAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACCA  
AGGAGAACGACACCATCATCTGCCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG  
GAGGTGGGCAAGGCCATGTACGCCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAA  
TATTACCGGCTGTGCTGACCCGCGACGGCGGCCACCAACAACAACCGCACCAACGACA  
CCGAGACCTTCCGCCCCGCGGCGGCCCAACATGAAGGACAACCTGGCGCAGCGAGCTGTAC  
AAGTACAAGGTGGTGCGCATCGAGCCCCCTGGGCGTGGCCCCCAACCAGGCCAAGCGCCG  
CGTGGTGACGCGGAGAAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTCATCGGCTTCCTGG  
GCGCCGCGCGGAGCACCATGGGCGCCGCTCCGTGACCCCTGACCGTGACGGCCCGCCAG  
CTGCTGAGCGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGAGGCCAGCA  
GCACCTGCTGACGCTGACCGTGTGGGCGATCAAGCAGCTGACGGCCCGCATCCTGGCCG  
TGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTG  
ATCTGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGAT  
CTGGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGA  
TCTACAACCTGATCGAGATCGCCGAGAACCAGCAGGAGAAGAACGAGCAGGAGCTGCTG  
GAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGGCTGTGGTA  
CATCCGCATCTTCATCATGATCGTGGGCGGCTGATCGGCCTGCGCATCGTGTTTCGCCG  
TGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCCTGCAGACCCGC  
CTGCCCCGCCAGCGCGGCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCG  
CGACCGCGACCGCAGCAACCGCCTGGTGACCGCCTGCTGGCCCTGATCTGGGACGACC  
TGCGCAGCCTGTGCCTGTTTCAGCTACCACCGCCTGCGCGACCTGCTGCTGATCGTGCC  
CGCATCGTGAGCTGCTGGGCGCGCGCTGGGAGGCCCTGAAGTACTGGTGGAACCT  
GCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCG  
CCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGATCGTGACGCGCATCTTCCGC  
GCCGTGATCCACATCCCCCGCCGCATCCGCCAGGGCCTGGAGCGCGCCCTGCTGTAAGA  
TATCGGATCCTCTAGA

FIG. 53

(SEQ ID NO:66)

SUBSTITUTE SHEET (RULE 26)

66 / 131

## gp160.modUS4delV1/2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCAACCGTGCTGTGGGTGACCGTGTACTACGGCGTGCCCGTG  
TGGAAGGAGGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAGGAGGTGAACC  
TGACCAACGTGACCGGAGAACTTCAACATGTGGAAGAAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCGCCGCGCCAGGCCTGCCC  
CAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTCGCCATCCTGAA  
GTGCAAGGACAAGAAGTTCAACGGCACCGGCCCTGCAAGAACGTGAGCACCGTGCAAGTGCA  
CCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAG  
GAGATCGTGCTGCGCTCCGAGAACTTACCGACAACGCCAAGACCATCATCGTGACAGCTGAA  
CGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAACACGCGTAAGAGCATCCACATCG  
GCCCCGCGCGCCTTCTACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCATCGA  
ACATCAGCAAGGCCAACTGGACCAACACCTCGAGCAGATCGTGGAAGAGCTGCGCGAGCAG  
TTCGGCAACAACAAGACCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGATCGTGTT  
CCACAGCTTCAACTGCGGCGGCGAGTTCTTCTACTGCAACACAGCCAGCTGTTCAACAGCAC  
CTGGAACATCACCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCC  
GCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATC  
CGCGGCCAGATCAAGTGACGAGCAATATTACCGCCTGCTGCTGACCCGCGACGGCGGCAC  
CAACAACAACCGCACCAACGACACCGAGACCTTCCGCCCCGGCGGCGCAACATGAAGGACA  
ACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGGCCCCC  
ACCCAGGCCAAGCGCCGCTGGTGACGCGGAGAAAGCGCGCCGTGGGCTGGGCGCCCTGTT  
CATCGGCTTCTGGGCGCCGCCGGGAGCACCATGGGCGCCGCTCCGTGACCTGACCGTGCA  
GGCCCGCCAGCTGCTGAGCGGCATCGTGACGAGCAGAGAACAACCTGCTGCGCGCCATCGAGG  
CCCAGCAGCACCTGCTGACGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTG  
GCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCT  
GATCTGCACCAACACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCT  
GGGACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTAC  
AACCTGATCGAGATCGCCCAGAACCCAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGG  
ACAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCACCAACTGGCTGTGGTACATCCGCATCT  
TCATCATGATCGTGGGCGGCCTGATCGGCCTGCGCATCGTGTTGCGCGTGCTGAGCATCGTGA  
ACCGCGTGCGCCAGGGCTACAGCCCCATCAGCCTGCAAGACCCGCTGCCCGCCCAGCGCGGC  
CCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCGCGACCGCGACCGCAGCAACC  
GCCTGGTGACGGCCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCCTGTTAGCT  
ACCACCGCCTGCGCGACCTGCTGCTGATCGTGCCCGCATCGTGGAGCTGCTGGGCGCGCGG  
GCTGGGAGGCCCTGAAGTACTGGTGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGC  
AGCGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCAT  
GAGATCGTGACGCGCATCTTCCGCGCCGTGATCCACATCCCCCGCCGCATCCGCCAGGGCCTG  
GAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA

FIG. 54

(SEQ ID NO:67)

SUBSTITUTE SHEET (RULE 26)

67 / 131

gp160.modUS4 del 128-194

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGAGCA  
GTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGCGCTGCCCCGTG  
TGGAAGGAGGCCACCAACCAACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAGGCCGAGGC  
CCACAACGTGTGGGCCACCAACGCTGCGTGCCACCGACCCCAACCCCAAGGAGGTGAACC  
TGACCAACGTGACCGAGAACCTCAACATGTGGAAGAACAACATGGTGGAGCAGATGCATGAG  
GACATCATCAGCCTGTGGGACAGAGCCTGAAGCCCTGCGTGAAGCTGACCCCCCTGTGCGTG  
GGGGCAGGGAACCTGCGAGACCAAGCGTGATCACCCAGGCTGCCCCAAGGTGAGCTTCGAGCC  
CATCCCCATCCACTACTGCGCCCCCGCGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTT  
CAACGGCACCCGGCCCCCTGCAAGAACGTGAGCACCCTGCAAGTGACCCACGGCATCCGCCCCG  
TGGTGAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCC  
GAGAACCTTACCGACAACGCCAAGACCATCATCGTGACGCTGAACGAGTCCGTTGGAGATCAA  
CTGCATCCGCCCCAACAACAACACGCGTAAGAGCATCCACATCGGCCCGGCCGCGCTTCTA  
CGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCACTGCAACATCAGCAAGGCCAACT  
GGACCAACACCCCTCGAGCAGATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACC  
ATCATCTTCAACAGCAGCAGCGCGGCGACCCCGAGATCGTGTTCACAGCTTCAACTGCGGC  
GGCGAGTTCTTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAG  
GTGAACAAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGATCCGCCAGATCATCAA  
CATGTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCA  
GCAGCAATATTACCGGCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCAAC  
GACACCGAGACCTTCCGCCCCGGCGGCGCAACATGAAGGACAACCTGGCGCAGCGAGCTGTA  
CAAGTACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGCCCCCACCAGGCCAAGCGCCGCG  
TGGTGACGCGGAGAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTCATCGGCTTCTTGGGCGCG  
CCGGGAGCACCATGGGCGCCGCTCCGTGACCCTGACCGTGACGGCCGCCAGCTGCTGAGC  
GGCATCGTGACGAGCAGAACAACCTGCTGCGCGCCATCGAGGCCAGCAGCACCTGCTGCA  
GCTGACCGTGTTGGGCGATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCGCTACCTGA  
AGGACCAGCAGCTGCTGGGCGATCTGGGGCTGCAGCGGCAAGCTGATCTGCACCACCACCGTG  
CCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAACATGACCTGGAT  
GGAGTGGGAGCGCGAGATCGGCAACTACACCGGCTGATCTACAACCTGATCGAGATCGCCC  
AGAACCAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTG  
GAACTGGTTTCGACATCAACCAACTGGCTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGG  
CCTGATCGGCCTGCGCATCGTGTTCCGCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTA  
CAGCCCCATCAGCCTGCAGACCCGCTGCCCCGCCAGCGCGGCCCGGACCGCCCCGAGGGCA  
TCGAGGAGGAGGGCGGCGAGCGCGACCGCAGCAACCGCCTGGTGACGGCCTGCTG  
GCCCTGATCTGGGACGACCTGCGCAGCCTGTGCCTGTTACGCTACCAACCGCCTGCGCGACCTG  
CTGCTGATCGTGGCCCGCATCGTGGAGCTGCTGGGCCCGCGGGCTGGGAGGCCCTGAAGTAC  
TGGTGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAA  
CGCCACCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGATCGTGACGCGCATCTT  
CCGCGCCGTGATCCACATCCCCCGCCGATCCGCCAGGGCCTGGAGCGCGCCCTGCTGTAAGA  
TATCGGATCCTCTAGA

FIG. 55

(SEQ ID NO:68)

SUBSTITUTE SHEET (RULE 26)

68 / 131

**Env\_US4\_C4wt**

GACACTATCATACTCCCATGCAGAATAAGACAAATTATAAACATGTGGCAAGAAGTAGG  
AAAAGCAATGTATGCCCCCTCCCATCAGAGGACAAATTAAATGTTTCATCAAATATTACAG  
GGCTGCTATTAAGTAGAGATGGTGGT

**FIG. 56**

(SEQ ID NO:69)

**Env\_SF162\_C4wt**

GGAACATATCACAATCCCATGCAGAATAAAACAAATTATAAACAGGTGGCAGGAAGTAGG  
AAAAGCAATGTATGCCCCCTCCCATCAGAGGACAAATTAGATGCTCATCAAATATTACAG  
GACTGCTATTAACAAGAGATGGTGGT

**FIG. 57**

(SEQ ID NO:70)

**Env\_US4\_C4mod**

GACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAGGAGGTGGG  
CAAGGCCATGTACGCCCCCCCCCATCCGCGGCCAGATCAAGTGCAGCAGCAACATCACCG  
GCCTGCTGCTGACCCGCGACGGCGGC

**FIG. 58**

(SEQ ID NO:71)

**Env\_SF162\_C4mod**

GGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGG  
CAAGGCCATGTACGCCCCCCCCCATCCGCGGCCAGATCCGCTGCAGCAGCAACATCACCG  
GCCTGCTGCTGACCCGCGACGGCGGC

**FIG. 59**

(SEQ ID NO:72)

*SUBSTITUTE SHEET (RULE 26)*

69 / 131

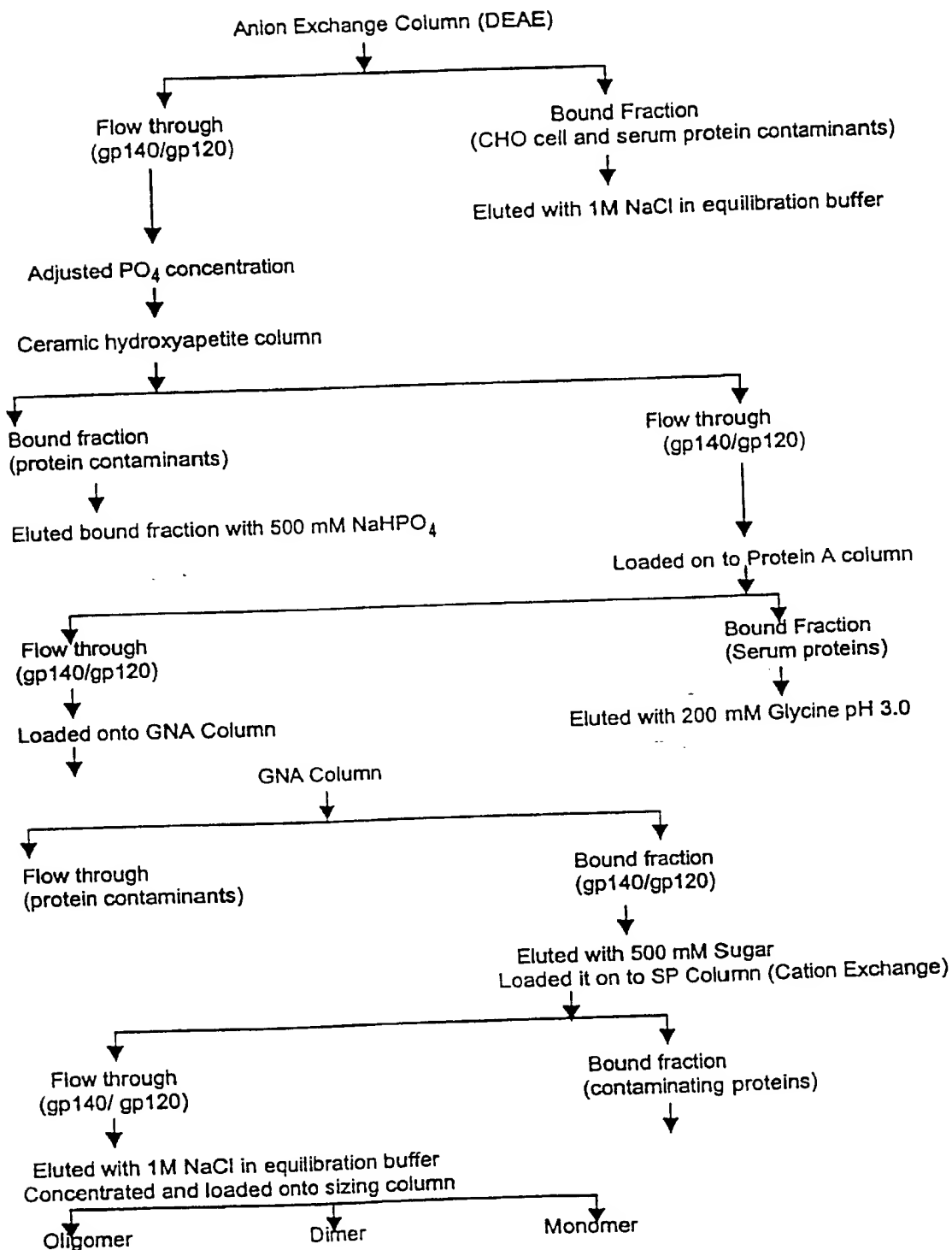


FIG. 60

SUBSTITUTE SHEET (RULE 26)

70 / 131

## gp160mod.us4.gag.modSF2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGA  
GCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGCGCTG  
CCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAG  
GCCGAGGCCCAACCGTGTGGGCCACCCACGCCTGCGTGCCCAACGACCCCAACCCCCAG  
GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG  
CAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG  
ACCCCCCTGTGCGTGACCTGAACTGCACCGACAAGCTGACCGGCAGCACCAACGGCACC  
AACAGCACCAGCGGCACCAACAGCACCAGCGGACCAACAGCACCAGCACCAACAGCACC  
GACAGCTGGGAGAAGATGCCCCGAGGGCAGATCAAGAAGTGCAGCTTCAACATCACCACC  
AGCGTGCGCGACAAGGTGCAGAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC  
ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCAACACCAGCGTGATCACCAG  
GCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTC  
GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGCCCCCTGCAAGAAGCTGAGC  
ACCGTGCAAGTGACCCAGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC  
AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTACCGACAACGCCAAGACC  
ATCATCGTGCAAGTGAACGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACAG  
CGTAAGAGCATCCACATCGGCCCGCGCGCCTTCTACGCCACCGGCGACATCATCGGC  
GACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG  
ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGACCATCATCTTCAACAGCAGC  
AGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC  
TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAACAAGACC  
AAGGAGAACGACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG  
GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAAGCAAT  
ATTACCGGCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC  
GAGACCTTCCGCCCCGCGCGCGCAACATGAAGGACAAGTGGCGCAGCGAGCTGTACAAG  
TACAAGGTGGTGCGCATCGAGCCCCTGGGCGTGCCCCCAACCGGCCAAGCGCCGCGTG  
GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGTTCTCGGCTTCTTGGGCGCC  
GCCGGGAGCACCATGGGCGCCGCTCCGTGACCCTGACCGTGCAAGGCCCGCCAGCTGCTG  
AGCGGCATCGTGCAAGCAGCAGAACAACCTGCTGCGCGCCATCGAGGCCCGAGCAGCACTG  
CTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTGGAGCGC  
TACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATCTGCACC  
ACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAAC  
ATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTACAACCTG  
ATCGAGATCGCCCAGAACCAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAAG  
TGGGCCAGCCTGTGGAAGTGGTTCGACATCAACCAACTGGCTGTGGTACATCCGCATCTTC  
ATCATGATCGTGCGCGCCCTGATCGGCCTGCGCATCGTGTTGCGCGTGCTGAGCATCGTG  
AACC GCGTGCGCCAGGGCTACAGCCCCATCAGCCTGCAGACCCGCTGCCCCGCCAGCGC  
GGCCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCGGACCGCGACCGCAGC  
AACC GCGTGGTGACGGCCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCCTGTGCCTG  
TTCAGCTACCACCGCCTGCGCGACCTGCTGCTGATCGTGCCCGCATCGTGAGCTGCTG  
GGCCGCGCGGCTGGGAGGCCCTGAAGTACTGGTGGAACCTGCTGCAGTACTGGAGCCAG  
GAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCCGAGGGC  
ACCGACCGCATCATCGAGATCGTGAGCGCATCTTCCGCGCCGTGATCCACATCCCCCGC  
CGCATCCGCCAGGGCCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGAGAATTC

FIG. 61A

(SEQ ID NO:73)

SUBSTITUTE SHEET (RULE 26)



71 / 131

CGCCCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGAAGCCGC  
TTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCGTCTTTT  
GGCAATGTGAGGGCCCCGAAACCTGGCCCTGTCTTCTTGACGAGCATTCTAGGGGTCTT  
TCCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTG  
GAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCCTTGCAGGCAGCGGAACCCCCCA  
CCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATAACCTGCAAAGGCG  
GCACAACCCCAAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGGCTCTCC  
TCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCATTGTATGGGATCT  
GATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAACGTCTA  
GGCCCCCGAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATAATACCATGGGCGC  
CCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGTGGGAGAAGATCCGCCTGCGCCC  
CGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGGCCAGCCGCGAGCTGGAGCG  
CTTCGCCGTGAACCCCGGCTGCTGGAGACCAGCGAGGGCTGCCGCCAGATCCTGGGCCA  
GCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGCGCAGCCTGTACAACACCGTGGC  
CACCTGTACTGCGTGACACAGCGCATCGACGTCAAGGACACCAAGGAGGCCCTGGAGAA  
GATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCAGCAGGCCGCCGCCGCCGCCG  
CACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCAGGGCCA  
GATGGTGCACACAGGCCATCAGCCCCCGCACCTGAACGCCTGGGTGAAGGTGGTGGAGGA  
GAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTACGCGCCCTGAGCGAGGGCGCCACCCC  
CCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCAGATGCT  
GAAGGAGACCATCAACGAGGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGCACGCCGG  
CCCCATCGCCCCCGGCCAGATGCGCGAGCCCCCGCGGCAGCGACATCGCCCGGCACCACCAG  
CACCTGTCAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGGCGAGAT  
CTACAAGCGGTGGATCATCCTGGGCCTGAACAAGATCGTGCGGATGTACAGCCCCACCAG  
CATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCGCTTCTA  
CAAGACCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAGAAGTGGATGACCGAGACCTT  
GCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCCTGAAGGCTCTCGGCCCGCGGGC  
CACCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCCGGCCACAAGGCCCG  
CGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCAGCGCGG  
CAACTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGGCCACAC  
CGCCAGGAAGTCCGCGCCCCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGAGGGCCA  
CCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCCCAGCTA  
CAAGGGCCGCCCCGGCAACTTCTGTCAGAGCCGCCCCGAGCCACCGCCCCCCCCGAGGA  
GAGCTTCCGCTTCGGCGAGGAGAAGACCCCCAGCCAGAAGCAGGAGCCCATCGACAA  
GGAGCTGTACCCCTGACCAGCCTGCGCAGCCTGTTTCGGCAACGACCCAGCAGCCAGTA  
AGAATTCACTCGAGCAAGTCTAGA

**FIG. 61B**

(SEQ ID NO:73)

*SUBSTITUTE SHEET (RULE 26)*

72 / 131

gp160mod.SF162.gag.modSF2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGG  
AGCAGTCTTCGTTTCGCCCAGCGCCGTGGAGAAGCTGTGGGTGACCGTGTACTACGGCG  
TGCCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTAC  
GACACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCC  
CCAGGAGATCGTGCTGGAGAACGTGACCGAGAATTCAACATGTGGAAGAACAACATGG  
TGGAGCAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTG  
AAGCTGACCCCCCTGTGCGTGACCTGCACTGCACCAACCTGAAGAACGCCACCAACAC  
CAAGAGCAGCAACTGGAAGGAGATGGACCGCGCGAGATCAAGAACTGCAGCTTCAAGG  
TGACCACCAGCATCCGCAACAGATGCAGAAGGAGTACGCCCTGTTCTACAAGCTGGAC  
GTGGTGCCCATCGACAACGACAACACCAGCTACAAGCTGATCAACTGCAACACCAGCGT  
GATCACCCAGGCCTGCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCC  
CCGCCGGCTTCGCCATCCTGAAGTGCAACGACAAGAAGTTCAACGGCAGCGGCCCTGCG  
ACCAACGTGAGCACCCTGTCAGTGACCCACGGCATCCGCCCCGTGGTGAGCACCAGCT  
GCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGTGGTGATCCGCAGCGAGAATTACCGG  
ACAACGCCAAGACCATCATCGTGACGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGC  
CCCAACAACAACCCCGCAAGAGCATCACCATCGGCCCGGCCGCGCCTTCTACGCCAC  
CGGCGACATCATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCGGCGAGAAGTGGA  
ACAACACCCTGAAGCAGATCGTGACCAAGCTGCAGGCCAGTTCCGCAACAAGACCATC  
GTGTTCAAGCAGAGCAGCGCGCGGACCCCGAGATCGTGATGCACAGCTTCAACTGCGG  
CGGCGAGTTCTTCTACTGCAACAGCACCCAGCTGTTCAACAGCACCTGGAACAACACCA  
TCGGCCCCAACAACACCAACGGCACCATCACCTGCCCTGCCGCATCAAGCAGATCATC  
AACCGCTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCCG  
CTGCAGCAGCAACATCACCGGCCTGCTGCTGACCCGCGACGGCGGCAAGGAGATCAGCA  
ACACCACCGAGATCTTCCGCCCCGGCGGCGGCGACATGCGCGACAACCTGGCGCAGCGAG  
CTGTACAAGTACAAGGTGGTGAAGATCGAGCCCCTGGGCGTGGCCCCCACCAGGCCAA  
GCGCCGCGTGGTGACGCGGAGAAGCGCGCCGTGACCTGGGCGCATGTTCTTGGGCT  
TCCTGGGCGCCGCCGCGCAGCACCATGGGCGCCCCGACCTGACCTGACCGTGACGGCC  
CGCCAGCTGCTGAGCGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGAGGC  
CCAGCAGCACCTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCCGCGTG  
TGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGC  
AAGCTGATCTGCACCACCGCCGTGCCCTGGAACGCCAGCTGGAGCAACAAGAGCCTGGA  
CCAGATCTGGAACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGACAACCTACACCA  
ACCTGATCTACACCCTGATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAGCAGGAG  
CTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAACCTGGTTTCGACATCAGCAAGTGGCT  
GTGGTACATCAAGATCTTCATCATGATCGTGGGCGGCCTGGTGGGCCTGCGCATCGTG  
TCACCGTGCTGAGCATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCCTGAGCTTCCAG  
ACCCGCTTCCCCGCCCCCGCGGCCCGACCCCGAGGGCATCGAGGAGGAGGGCGG  
CGAGCGGACCGCGACCGCAGCAGCCCCCTGGTGACGGCCTGCTGGCCCTGATCTGGG  
ACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCGCCTGCGCGACCTGATCCTGATC  
GCCGCCCGCATCGTGAGCTGCTGGGCCGCGCGGCTGGGAGGCCCTGAAGTACTGGGG  
CAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGCGCCGTGAGCCTGTTTCGACG  
CCATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCCAGCGCATC  
GGCCGCGCCTTCTGACATCCCCCGCCGCATCCGCCAGGGCTTCGAGCGCGCCCTGCT

FIG. 62A

(SEQ ID NO:74)

SUBSTITUTE SHEET (RULE 26)



74 / 131

gp160modUS4.delV1/V2.gag.modSF2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCTGCTGTGTGGA  
GCAGTCTTCGTTTCGCCCAGCGCCACCACCGTGCTGTGGGTGACCGTGTAACGCGTG  
CCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCTTACAAG  
GCCGAGGCCCACAACTGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG  
GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG  
CAGATGCATGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGGGCGCC  
GGCCAGGCCTGCCCCAAGGTGAGCTTCGAGCCCATCCCCATCCACTACTGCGCCCCCGCC  
GGCTTCGCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCGGCCCCCTGCAAGAAC  
GTGAGCACCGTGCACTGCACCCACGGCATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTG  
AACGGCAGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACTTCACCGACAACGCC  
AAGACCATCATCGTGCACTGAAAGAGTCCGTGGAGATCAACTGCATCCGCCCCAACAAAC  
AACACGCGTAAGAGCATCCACATCGGCCCCGGCGCGCCTTCTACGCCACCGGCGACATC  
ATCGGCGACATCCGCCAGGCCCACTGCAACATCAGCAAGGCCAACTGGACCAACACCCTC  
GAGCAGATCGTGGAAGAAGCTGCGCGAGCAGTTCCGGCAACAACAAGACCATCATCTTCAAC  
AGCAGCAGCGGCGGCGACCCCGAGATCGTGTTCCACAGCTTCAACTGCGGCGGGCGAGTTC  
TTCTACTGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCACCGAGGAGGTGAAC  
AAGACCAAGGAGAACGACACCATCATCCTGCCCTGCCGCATCCGCCAGATCATCAACATG  
TGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCCAGATCAAGTGCAGC  
AGCAATATTACCGGCCCTGCTGCTGACCCGCGACGGCGGCACCAACAACAACCGCACCAAC  
GACACCGAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAACCTGGCGCAGCGAGCTG  
TACAAGTACAAGGTGGTGCATCGAGCCCCCTGGGCGTGGCCCCCACCCAGGCCAAGCGC  
CGCGTGGTGCAGCGCGAGAAGCGCGCCGTGGGCGCTGGGCGCCCTGTTTCATCGGCTTCCTG  
GGCGCCGCCGGGAGCACCATGGGCGCCGCTCCGTGACCCCTGACCGTGACGGCCCCGCCAG  
CTGCTGAGCGGCATCGTGACGAGCAGACAACCTGCTGCGCGCCATCGAGGCCCCAGCAG  
CACCTGCTGCAGCTGACCGTGTGGGGCATCAAGCAGCTGCAGGCCCGCATCCTGGCCGTG  
GAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGCAGCGGCAAGCTGATC  
TGCACCACCACCGTGCCCTGGAACAGCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGG  
GACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTACACCGGCCTGATCTAC  
AACCTGATCGAGATCGCCAGAACAGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTG  
GACAAGTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGGCTGTGGTACATCCGC  
ATCTTCATCATGATCGTGGGCGGCCCTGATCGGCCTGCGCATCGTCTTCGCCGTGCTGAGC  
ATCGTGAACCGCGTGCGCCAGGGCTACAGCCCCATCAGCCTGCAGACCCGCCTGCCCGCC  
CAGCGCGGCCCCGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGAGCGCGACCGCGAC  
CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTGCGCAGCCTG  
TGCTTGTTCAGCTACCACCGCCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG  
CTGCTGGGCGCCCGCGGCTGGGAGGCCCTGAAGTACTGGTGGAACCTGCTGCAGTACTGG  
AGCCAGGAGCTGAAGAGCAGCGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC  
GAGGGCACCGACCGCATCATCGAGATCGTGACGCGCATCTTCCGCGCCGTGATCCACATC  
CCCCGCCGCATCCGCCAGGGCCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA  
GAATTCGCCCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCCTAACGTTACTGGCCGA  
AGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATTTTCCACCATATTGCCG  
TCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCCTAGG  
GGTCTTTCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTT

FIG. 63A

(SEQ ID NO:75)

SUBSTITUTE SHEET (RULE 26)

75 / 131

CCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAAC  
CCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGCA  
AAGGCGGCACAACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAAGAGTCAAATGG  
CTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATTTGTATG  
GGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAA  
CGTCTAGGCCCCCGAACCACGGGGACGTGGTTTTCTTTGAAAAACACGATAATACCAT  
GGGCGCCCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGTGGGAGAAGATCCGCCT  
GCGCCCCGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGGGCCAGCCGCGAGCT  
GGAGCGCTTCGCCGTGAACCCCGGCCTGCTGGAGACCAGCGAGGGCTGCCGCCAGATCCT  
GGGCCAGCTGCAGCCAGCCTGCAGACCGGCAGCGAGGAGCTGCGCAGCCTGTACAACAC  
CGTGGCCACCCTGTACTGCGTGACCCAGCGCATCGACGTCAAGGACACCAAGGAGGCCCT  
GGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCAGCAGGCCGCCGCCGC  
CGCCGGCACCAGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATCGTGCAGAACCTGCA  
GGGCCAGATGGTGCACCCAGGCCATCAGCCCCCGCACCCCTGAACGCCTGGGTGAAGGTGGT  
GGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTACGCGCCCTGAGCGAGGGCGC  
CACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCACCAGGCCGCCATGCA  
GATGCTGAAGGAGACCATCAACGAGGAGGCCGCCGAGTGGGACCGCGTGCACCCCGTGCA  
CGCCGGCCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGCGACATCGCCGGCAC  
CACCAGCACCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCCCCATCCCCGTGGG  
CGAGATCTACAAGCGGTGGATCATCTGGGCCTGAACAAGATCGTGCGGATGTACAGCCC  
CACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCGCGACTACGTGGACCG  
CTTCTACAAGACCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAGAACTGGATGACCGA  
GACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCCTGAAGGCTCTCGGCCC  
CGCGGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGGGCGGCCCGGCCACAA  
GGCCCGCGTGCTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCGACCATCATGATGCA  
GCGCGGCAACTTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAACTGCGGCAAGGAGGG  
CCACACCGCCAGGAAGTCCCGCGCCCCCGCAAGAAGGGCTGCTGGCGCTGCGGCCGCGA  
GGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCTGGGCAAGATCTGGCC  
CAGCTACAAGGGCCGCCCGGCAACTTCTGCGAGCGCCCGGAGCCACCGCCCCCCC  
CGAGGAGAGCTTCCGCTTCGGCGAGGAGAAGACCACCCCGCCAGCAGAAGCAGGAGCCCAT  
CGACAAGGAGCTGTACCCCTGACCAGCCTGCGCAGCCTGTTGGCAACGACCCCGAGCAG  
CCAGTAAGAATTCAGACTCGAGCAAGTCTAGA

**FIG. 63B**

(SEQ ID NO:75)

*SUBSTITUTE SHEET (RULE 26)*

76 / 131

## gp160.modSF162.delV2.gag.modSF2

GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCTGTGTGCTGCTGCTGTGTGGA  
GCAGTCTTCGTTTCGCCCAGCGCCGTGGAGAAGCTGTGGGTGACCGTGTAACGCGCTG  
CCCGTGTGGAAGGAGGCCACCACCACCTGTTCTGCGCCAGCGACGCCAAGGCCTACGAC  
ACCGAGGTGCACAACGTGTGGGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG  
GAGATCGTGCTGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG  
CAGATGCACGAGGACATCATCAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG  
ACCCCCCTGTGCGTGACCTGCACTGCACCAACCTGAAGAACGCCACCAACACCAAGAGC  
AGCAACTGGAAGGAGATGGACCGCGGCGAGATCAAGAACTGCAGCTTCAAGGTGGGCGCC  
GGCAAGCTGATCAACTGCAACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTTC  
GAGCCCATCCCCATCCACTACTGCGCCCCCGCGGCTTCGCCATCCTGAAGTGCAACGAC  
AAGAAGTTCAACGGCAGCGGCCCTGCACCAACGTGAGCACCCTGCGTGCAGTGCACCCACGGC  
ATCCGCCCCGTGGTGAGCACCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGTG  
GTGATCCGCGAGCGAGAAGTTCAACGACAACGCCAAGACCATCATCGTGAGCTGAAGGAG  
AGCGTGAGATCAACTGCACCCGCCCCAACAACAACACCCGCAAGAGCATCACCATCGGC  
CCCCGCGCGCCTTCTACGCCACCGGCGACATCATCGGCGACATCCGCCAGGCCCCACTGC  
AACATCAGCGGCGAGAAGTGGAACAACACCCTGAAGCAGATCGTGACCAAGCTGCAGGCC  
CAGTTCGGCAACAAGACCATCGTGTTCAAGCAGAGCAGCGGCGGCGACCCCGAGATCGTG  
ATGCACAGCTTCAACTGCGGCGGCGAGTTCTTCTACTGCAACAGCACCCAGCTGTTCAAC  
AGCACCTGGAACAACACCATCGGCCCCAACAACACCAACGGCACCATCACCTGCCCTGC  
CGCATCAAGCAGATCATCAACCGCTGGCAGGAGGTGGGCAAGGCCATGTACGCCCCCCCC  
ATCCGCGGCCAGATCCGCTGCAGCAGCAACATCACCGGCCTGCTGCTGACCCGCGACGGC  
GGCAAGGAGATCAGCAACACCACCGAGATCTTCCGCCCCGCGGCGGCGACATGCGCGAC  
AACTGGCGCAGCGAGCTGTACAAGTACAAGGTGGTGAAGATCGAGCCCCCTGGGCGTGGCC  
CCCACCAAGGCCAAGCGCCGCGTGGTGACGCGCGAGAAGCGCGCCGTGACCCCTGGGCGCC  
ATGTTCTTGGGCTTCTTGGGCGCGCGCGGCGAGCACCATGGGCGCCCCGAGCCTGACCCTG  
ACCGTGACGGCCCGCCAGCTGCTGAGCGGCATCGTGACGAGCAGAGAACAACCTGCTGCGC  
GCCATCGAGGCCCGCAGCAGCACCCTGCTGCAGCTGACCGTGTTGGGCGATCAAGCAGCTGCAG  
GCCCCGCTGCTGGCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGGGCGATCTGGGCG  
TGCAGCGGCAAGCTGATCTGCACCAACCGCCGTGCCCTGGAACGCCAGCTGGAGCAACAAG  
AGCCTGGACCAGATCTGGAACAACATGACCTGGATGGAGTGGGAGCGCGAGATCGACAAC  
TACACCAACCTGATCTACACCCTGATCGAGGAGAGCCAGAACCAGCAGGAGAAGAACGAG  
CAGGAGCTGCTGGAGCTGGACAAGTGGGCCAGCCTGTGGAAGTGGTTCGACATCAGCAAG  
TGGCTGTGGTACATCAAGATCTTCATCATGATCGTGCGGCGGCTGAGCCCCCTGAGCTTC  
GTGTTACCGTGCTGAGCATCGTGAACCGCGTGCAGCGGCTACAGCCCCCTGAGCTTC  
CAGACCCGCTTCCCCGCCCCCGCGCGGCCGAGGCCATCGAGGAGGAGGGC  
GGCGAGCGCGACCGCAGCAGCCCCCTGGTGCACGGCCTGCTGGCCCTGATCTGG  
GACGACCTGCGCAGCCTGTGCCTGTTAGCTACCAACCGCCTGCGCGACCTGATCCTGATC  
GCCGCCCCGATCGTGAGCTGCTGGGCCGCGCGGCTGGGAGGCCCTGAAGTACTGGGGC  
AACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGCGCCGTGAGCCTGTTTCGACGCC  
ATCGCCATCGCCGTGGCCGAGGGCACCGACCGCATCATCGAGGTGGCCAGCGCATCGGC  
CGCGCCTTCTTGACATCCCCCGCGCATCCGCCAGGGCTTCGAGCGCGCCCTGCTGTAA  
CTCGAGCAAGTCTAGAGAATTCCGCCCCCCCCCCCCCCCCCTCTCCCTCCCCCCCCC  
TAACGTTACTGGCCGAAGCGCTTGAATAAGGCCGGTGTGCGTTTGTCTATATGTTATT  
TTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGAAACCTGGCCCTGTCTTCTT

FIG. 64A

(SEQ ID NO:76)

SUBSTITUTE SHEET (RULE 26)

77 / 131

GACGAGCATTTCCTAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGT  
CGTGAAGGAAGCAGTTTCCTCTGGAAGCTTCTTGAAGACAAACAACGTCTGTAGCGACCCCT  
TTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGT  
ATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAGTTGT  
GGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAA  
GGTACCCCATTTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTA  
GTCGAGGTTAAAAAACGTCTAGGCCCCCGAACCACGGGGACGTGGTTTTCTTTGAAA  
AACACGATAATACCATGGGCGCCCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGT  
GGGAGAAGATCCGCCTGCGCCCCGCGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGT  
GGGCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCCTGCTGGAGACCAGCGAGG  
GCTGCCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGC  
GCAGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGCACCAGCGCATCGACGTCAAGG  
ACACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCC  
AGCAGGCCCGCCGCCCGCCCGGCCACCGGCAACAGCAGCCAGGTGAGCCAGAATAACCCCA  
TCGTGCAGAACCTGCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCCCTGAACG  
CCTGGGTGAAGGTGGTGGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTTCAGCG  
CCCTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCC  
ACCAGGCCCGCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCCGCCGAGTGGGACC  
GCGTGCACCCCGTGCACGCGCGGCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCA  
GCGACATCGCCGGCACCACCAGCACCTGCAGGAGCAGATCGGCTGGATGACCAACAACC  
CCCCCATCCCCGTGGGCGAGATCTACAAGCGGTGGATCATCCTGGGCCTGAACAAGATCG  
TGCGGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCC  
GCGACTACGTGGACCGCTTCTACAAGACCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGA  
AGAAGGCTCTCGGCCCCCGCGGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGG  
TGAAGGCTCTCGGCCCCCGCGGCCACCCTGGAGGAGATGATGACCGCCTGCCAGGGCGTGG  
GCGGCCCCCGGCCACAAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGG  
CGACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCA  
ACTGCGGCAAGGAGGGCCACCGCCAGGAACTGCCGCGCCCCCGCAAGAAGGGCTGCT  
GGCGCTGCGGCCGCGAGGGCCACCAGATGAAGGACTGCACCGAGCGCCAGGCCAACTTCC  
TGGGCAAGATCTGGCCCAGCTACAAGGGCCGCCCGGCAACTTCTGCAGAGCCGCCCGG  
AGCCCACCGCCCCCCCCGAGGAGGCTTCCGCTTCGGCGAGGAGAAGACCACCCCCAGCC  
AGAAGCAGGAGCCCATCGACAAGGAGCTGTACCCCTGACCAGCCTGCGCAGCCTGTTTCG  
GCAACGACCCCGAGCAGCCAGTAAGAATTGAGACTCGAGCAAGTCTAGA

**FIG. 64B**

(SEQ ID NO:76)

*SUBSTITUTE SHEET (RULE 26)*

78 / 131

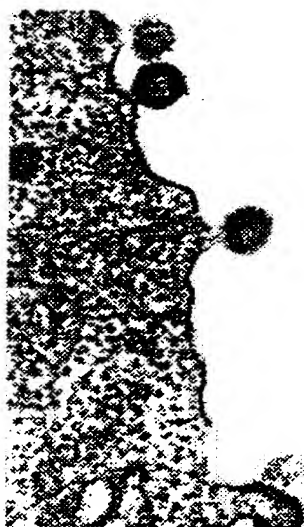


FIG. 65C



FIG. 65B



FIG. 65A



FIG. 65F



FIG. 65E



FIG. 65D

SUBSTITUTE SHEET (RULE 26)



79 / 131

gp160.modSF162	1	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	50
gp160.modSF162.delV2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp160.modSF162.delV1V2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.mut.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.mut7.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp140.mut8.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	
gp120.modSF162	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGCTCTGCTGTGTGCTGCT	100
Consensus	51	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp160.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp160.modSF162.delV2	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp160.modSF162.delV1V2	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp140.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp140.mut.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp140.mut7.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp140.mut8.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	
gp120.modSF162	(51)	GCTGTGTGGAGCAGTCTTCGTTTCGCCACGCCCGTGGAGAAGCTGTGGG	150
Consensus	101	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp160.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp160.modSF162.delV2	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp160.modSF162.delV1V2	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp140.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp140.mut.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp140.mut7.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp140.mut8.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
gp120.modSF162	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	
Consensus	(101)	TGACCGTGTACTACGGCGTGCCCGTGTGGAAGGAGGCCACCCACCCCTG	

FIG. 66A-1

SUBSTITUTE SHEET (RULE 26)

80 / 131

		151	200
gp160.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp160.modSF162.delV2	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp160.modSF162.delV1V2	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp140.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp140.mut.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp140.mut7.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp140.mut8.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
gp120.modSF162	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
Consensus	(151)	TTCTGGCCAGCGACGCCAAGGCCTACGACACCGAGGTGCACAACGTTG	
		201	250
gp160.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp160.modSF162.delV2	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp160.modSF162.delV1V2	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp140.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp140.mut.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp140.mut7.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp140.mut8.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
gp120.modSF162	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
Consensus	(201)	GGCCACCCACGCCTGCGTGCCCCACCGACCCCCAACCCCCAGGAGATCGTGC	
		251	300
gp160.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp160.modSF162.delV2	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp160.modSF162.delV1V2	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp140.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp140.mut.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp140.mut7.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	
gp140.mut8.modSF162	(251)	TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAACATGGTGGAG	

FIG. 66A-2

SUBSTITUTE SHEET (RULE 26)

81 / 131

gp120.modSF162  
Consensus  
(251) TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAG  
(251) TGGAGAACGTGACCGAGAACTTCAACATGTGGAAGAACAAACATGGTGGAG 350

gp160.modSF162  
gp160.modSF162.delV2  
gp160.modSF162.delV1V2  
gp140.modSF162  
gp140.mut.modSF162  
gp140.mut7.modSF162  
gp140.mut8.modSF162  
gp120.modSF162  
Consensus  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG  
(301) CAGATGCACGAGGACATCATCAGCCTGTGGACCAGAGCCTGAAGCCCTG 400

gp160.modSF162  
gp160.modSF162.delV2  
gp160.modSF162.delV1V2  
gp140.modSF162  
gp140.mut.modSF162  
gp140.mut7.modSF162  
gp140.mut8.modSF162  
gp120.modSF162  
Consensus  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA  
(351) CGTGAAGCTGACCCCCCTGTGCGTGACCCCTGCACCTGCACCAACCTGAAGA 450

gp160.modSF162  
gp160.modSF162.delV2  
gp160.modSF162.delV1V2  
gp140.modSF162  
gp140.mut.modSF162  
gp140.mut7.modSF162  
gp140.mut8.modSF162  
gp120.modSF162  
Consensus  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(375) -----  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG  
(401) ACGCCACCAACACCAAGAGCAGCAACTGGAAGGAGATGGACCGGGCGAG 450

FIG. 66A-3

SUBSTITUTE SHEET (RULE 26)

82 / 131

gp160.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	500
gp160.modSF162.delV2	(451)	ATCAAGAACTGCAGCTTCAAGGTGGC-----	
gp160.modSF162.delV1V2	(376)	-----GGC-----	
gp140.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
gp140.mut.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
gp140.mut7.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
gp140.mut8.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
gp120.modSF162	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	
Consensus	(451)	ATCAAGAACTGCAGCTTCAAGGTGACCAACAGCATCCGCAACAAGATGCA	550
gp160.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
gp160.modSF162.delV2	(478)	-----GCC-----GG-----	
gp160.modSF162.delV1V2	(379)	-----GCC-----GG-----	
gp140.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
gp140.mut.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
gp140.mut7.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
gp140.mut8.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
gp120.modSF162	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	
Consensus	(501)	GAAGGAGTACGCCCTGTTCTACAAGCTGGACGTGGTGCCCATCGACAACG	600
gp160.modSF162	(551)	ACACACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
gp160.modSF162.delV2	(492)	-----CAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
gp160.modSF162.delV1V2	(384)	-----CAACTGCCAGACCCAGCGTGATCACCCAG	
gp140.modSF162	(551)	ACAACACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
gp140.mut.modSF162	(551)	ACAACACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	

FIG. 66A-4

SUBSTITUTE SHEET (RULE 26)

83 / 131

gp140.mut7.modSF162	(551)	ACAAACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	650
gp140.mut8.modSF162	(551)	ACAAACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
gp120.modSF162	(551)	ACAAACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
Consensus	(551)	ACAAACACAGCTACAAGCTGATCAACTGCAACACACAGCGTGATCACCCAG	
gp160.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	601
gp160.modSF162.delV2	(520)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp160.modSF162.delV1V2	(412)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp140.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp140.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp140.mut.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp140.mut7.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp140.mut8.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp120.modSF162	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
Consensus	(601)	GCCTGCCCCAAGGTGAGCTTCGAGGCCATCCCCATCCCACTACTGCGCCCC	
gp160.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	651
gp160.modSF162.delV2	(570)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp160.modSF162.delV1V2	(462)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp140.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp140.mut.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp140.mut7.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp140.mut8.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp120.modSF162	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
Consensus	(651)	CGCCGGCTTCGCCATCCTGAAGTGCAACGACAAAGAGTTCAACGGCAGCG	
gp160.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	701
gp160.modSF162.delV2	(620)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp160.modSF162.delV1V2	(512)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp140.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp140.mut.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp140.mut7.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp140.mut8.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
gp120.modSF162	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	
Consensus	(701)	GCCCTGCACCAACGTGAGCACCGTGAGTGCAACCCACGGCATCCGCCCC	

FIG. 66A-5

SUBSTITUTE SHEET (RULE 26)

	751	800
gp160.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCCGAGGAGGGCGT
gp160.modSF162.delV2	(670)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp160.modSF162.delV1V2	(562)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp140.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp140.mut.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp140.mut7.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp140.mut8.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
gp120.modSF162	(751)	GTGTTAGCACCCAGCTGCTGCTGAACGGCAGCCTGGCCGAGGAGGGCGT
Consensus	(751)	850
gp160.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp160.modSF162.delV2	(720)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp160.modSF162.delV1V2	(612)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp140.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp140.mut.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp140.mut7.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp140.mut8.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
gp120.modSF162	(801)	GGTGATCCGCAGCGAGAACTTCACCGACAAACGCCAAGACCATCATCTGTCG
Consensus	(801)	900
gp160.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCCCCAACAACAACACC
gp160.modSF162.delV2	(770)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCCCCAACAACAACACC
gp160.modSF162.delV1V2	(662)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCCCCCAACAACAACACC
gp160.modSF162		

**FIG. 66A-6**

85 / 131

gp140.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCGCCCAACAACAACACC	950
gp140.mut.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCGCCCAACAACAACACC	
gp140.mut7.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCGCCCAACAACAACACC	
gp140.mut8.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCGCCCAACAACAACACC	
gp120.modSF162	(851)	AGCTGAAGGAGAGCGTGGAGATCAACTGCACCCGCGCCCAACAACAACACC	950
Consensus			
gp160.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	951
gp160.modSF162.delV2	(820)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp160.modSF162.delV1V2	(712)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp140.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp140.mut.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp140.mut7.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp140.mut8.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	
gp120.modSF162	(901)	CGCAAGAGCATCACCATCGGCCCCCGCGCGCTTCTACGCCACCGGCGGA	1000
Consensus			
gp160.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	951
gp160.modSF162.delV2	(870)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp160.modSF162.delV1V2	(762)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp140.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp140.mut.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp140.mut7.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp140.mut8.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
gp120.modSF162	(951)	CATCATCGGCGGACATCCGCCAGGCGCCACTGCAACATCAGCGGCGGAGAAGT	
Consensus			

FIG. 66A-7

SUBSTITUTE SHEET (RULE 26)

86 / 131

gp160.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	1050
gp160.modSF162.delV2	(920)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp160.modSF162.delV1V2	(812)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp140.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp140.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp140.mut.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp140.mut7.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp140.mut8.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
gp120.modSF162	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	
Consensus	(1001)	GGAACAACACCCCTGAAGCAGATCGTGACCAAGCTGCAGGGCCAGTTCGGC	1100
gp160.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	1051
gp160.modSF162.delV2	(970)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp160.modSF162.delV1V2	(862)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp140.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp140.mut.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp140.mut7.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp140.mut8.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
gp120.modSF162	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	
Consensus	(1051)	AACAAGACCATCGTGTCAAGCAGAGCAGCGGGCGGACCCCGAGATCGT	1150
gp160.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	1101
gp160.modSF162.delV2	(1020)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp160.modSF162.delV1V2	(912)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp140.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp140.mut.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp140.mut7.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp140.mut8.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
gp120.modSF162	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	
Consensus	(1101)	GATGCACAGCTTCAACTGCGGGCGGAGTTCTTCTACTGCAACAGCACCC	1200
gp160.modSF162	(1151)	AGCTGTTCACACAGCACCTGGAACAACACCATCGGGCCCCAACACCAAC	1151

FIG. 66A-8

SUBSTITUTE SHEET (RULE 26)



87 / 131

gp160.modSF162.delV2	(1070)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	1250
gp160.modSF162.delV1V2	(962)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp140.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
gp120.modSF162	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	
Consensus	(1151)	AGCTGTTCAACAGCACCTGGAACAACACCATCGGCCCAACAACACCAAC	1250
gp160.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	1201
gp160.modSF162.delV2	(1120)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp160.modSF162.delV1V2	(1012)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp140.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
gp120.modSF162	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	
Consensus	(1201)	GGCACCATCACCCCTGCCCTGCCGATCAAGCAGATCATCAACCGCTGGCA	1300
gp160.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	1251
gp160.modSF162.delV2	(1170)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp160.modSF162.delV1V2	(1062)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp140.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp140.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp140.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp140.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp140.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
gp120.modSF162	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	
Consensus	(1251)	GGAGGTGGCAAGGCCATGTACGCCCCCCCCCATCCGGGCCAGATCCGCT	1300

FIG. 66A-9

SUBSTITUTE SHEET (RULE 26)

88 / 131

gp160.modSF162	(1301)	1301	GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	1350
gp160.modSF162.delV2	(1220)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp160.modSF162.delV1V2	(1112)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp140.modSF162	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp140.mut.modSF162	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp140.mut7.modSF162	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp140.mut8.modSF162	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
gp120.modSF162	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	
Consensus	(1301)		GCAGCAGCAACATCACCGGCTGCTGCTGACCCGCGACGGCGGCAAGGAG	1400
gp160.modSF162	(1351)	1351	ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp160.modSF162.delV2	(1270)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp160.modSF162.delV1V2	(1162)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp140.modSF162	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp140.mut.modSF162	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp140.mut7.modSF162	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp140.mut8.modSF162	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
gp120.modSF162	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	
Consensus	(1351)		ATCAGCAACACACCGAGATCTTCCGCCCGCGCGCGGCGACATGCGCGA	1401
gp160.modSF162	(1401)	1401	CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp160.modSF162.delV2	(1320)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp160.modSF162.delV1V2	(1212)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp140.modSF162	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp140.mut.modSF162	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp140.mut7.modSF162	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp140.mut8.modSF162	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
gp120.modSF162	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	
Consensus	(1401)		CAACTGGCGCAGCGAGCTGTACAAGTACAAGTGGTGAAGATCGAGCCCC	1450

FIG. 66A-10

SUBSTITUTE SHEET (RULE 26)

89 / 131

gp160.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	1500
gp160.modSF162.delV2	(1370)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp160.modSF162.delV1V2	(1262)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp140.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp140.mut.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp140.mut7.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp140.mut8.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
gp120.modSF162	(1451)	TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	
Consensus		TGGCGTGGCCCCCACCAGGCCAAGCGCCGCTGCTGGCGCGGAGAAAG	1550
gp160.modSF162	(1501)	CGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	1501
gp160.modSF162.delV2	(1420)	CGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp160.modSF162.delV1V2	(1312)	CGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp140.modSF162	(1501)	CGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp140.mut.modSF162	(1501)	AGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp140.mut7.modSF162	(1501)	AGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp140.mut8.modSF162	(1501)	AGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	
gp120.modSF162	(1501)	CGC-----TAACTCGAG-----	
Consensus		CGCGCGTGACCCCTGGCGGCCATGTTCTGGGCTTCCTGGGCGCGCCGG	1600
gp160.modSF162	(1551)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	1551
gp160.modSF162.delV2	(1470)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp160.modSF162.delV1V2	(1362)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp140.modSF162	(1551)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp140.mut.modSF162	(1551)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp140.mut7.modSF162	(1551)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp140.mut8.modSF162	(1551)	CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	
gp120.modSF162	(1513)	-----	
Consensus		CAGCACCATGGGCGCCCGCAGCCTGACCCCTGACCGTGACGGCCCGCCAGC	

FIG. 66A-11

SUBSTITUTE SHEET (RULE 26)

90 / 131

1601	1650	
1601	1601	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp160.modSF162	(1601)	
gp160.modSF162.delV2	(1520)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp160.modSF162.delV1V2	(1412)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp140.modSF162	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp140.modSF162	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp140.mut.modSF162	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp140.mut7.modSF162	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp140.mut8.modSF162	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
gp120.modSF162	(1513)	-----
Consensus	(1601)	TGCTGAGCGGCATCGTGCAGCAGCAGAGAACAACCTGCTGCGGCCATCGAG
	1651	1700
gp160.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp160.modSF162.delV2	(1570)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp160.modSF162.delV1V2	(1462)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp140.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp140.mut.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp140.mut7.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp140.mut8.modSF162	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
gp120.modSF162	(1513)	-----
Consensus	(1651)	GCCAGCAGCACCTGCTGCAGCTGACCGTGCGGGGCATCAAGCAGCTGCA
	1701	1750
gp160.modSF162	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp160.modSF162.delV2	(1620)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp160.modSF162.delV1V2	(1512)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp140.modSF162	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp140.mut.modSF162	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp140.mut7.modSF162	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
gp140.mut8.modSF162	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG
	(1701)	GGCCCGGTGCTGGCCCGTGGAGCGCTACCTGAAGGACCAGCAGCTGCTGG

FIG. 66A-12

SUBSTITUTE SHEET (RULE 26)

91 / 131

gp120.modSF162	(1513)	-----
Consensus	(1701)	GGCCCGCTGCTGGCCCTGGAGCGCTACCTGAAGGACCAGCAGTGTGG
		1800
gp160.modSF162	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp160.modSF162.delV2	(1670)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp160.modSF162.delV1V2	(1562)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp140.modSF162	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp140.mut.modSF162	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp140.mut7.modSF162	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp140.mut8.modSF162	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
gp120.modSF162	(1513)	-----
Consensus	(1751)	GCATCTGGGGTGCAGCGGCAAGCTGATCTGCACCACCGCCGTGCCCTGG
		1801
gp160.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp160.modSF162.delV2	(1720)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp160.modSF162.delV1V2	(1612)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp140.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp140.mut.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp140.mut7.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp140.mut8.modSF162	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
gp120.modSF162	(1513)	-----
Consensus	(1801)	AACGCCAGCTGGAGCAACAAGAGCCCTGGACCAGATCTGGAACAACATGAC
		1851
gp160.modSF162	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp160.modSF162.delV2	(1770)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp160.modSF162.delV1V2	(1662)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp140.modSF162	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp140.mut.modSF162	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp140.mut7.modSF162	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp140.mut8.modSF162	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA
gp120.modSF162	(1513)	-----
Consensus	(1851)	CTGGATGGAGTGGGAGCGCGAGATCGACAACCTACCAACCTGATCTACA

FIG. 66A-13

SUBSTITUTE SHEET (RULE 26)

92/131

			1950
			1951
gp160.modSF162	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp160.modSF162.delV2	(1820)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp160.modSF162.delV1V2	(1712)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp140.modSF162	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp140.mut.modSF162	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp140.mut7.modSF162	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp140.mut8.modSF162	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
gp120.modSF162	(1513)	-----	
Consensus	(1901)	CCCTGATCGAGGAGAGCCAGAACCCAGCAGGAGGAGAAACGAGCAGGAGCTG	
			2000
			1951
gp160.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp160.modSF162.delV2	(1870)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp160.modSF162.delV1V2	(1762)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp140.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp140.mut.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp140.mut7.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp140.mut8.modSF162	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
gp120.modSF162	(1513)	-----	
Consensus	(1951)	CTGGAGCTGGACAAAGTGGGCCAGCCTGTGGAACCTGGTTCGACATCAGCAA	
			2050
gp160.modSF162	(2001)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGCGGCCCTGGTGG	
gp160.modSF162.delV2	(1920)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGCGGCCCTGGTGG	
gp160.modSF162.delV1V2	(1812)	GTGGCTGTGGTACATCAAGATCTTCATCATGATCGTGGCGGCCCTGGTGG	
gp140.modSF162	(2001)	GTGGCTGTGGTACATCTAACTCGAG-----	
gp140.mut.modSF162	(2001)	GTGGCTGTGGTACATCTAACTCGAG-----	

FIG. 66A-14

SUBSTITUTE SHEET (RULE 26)

93 / 131

gp140.mut7.modSF162	(2001)	GTGGCTGTGTACATCTAACTCGAG	2100
gp140.mut8.modSF162	(2001)	GTGGCTGTGTACATCTAACTCGAG	
gp120.modSF162	(1513)		
Consensus	(2001)	GTGGCTGTGTACATCTAACTCGAG	
gp160.modSF162	(2051)	GCCTGGGCATCGTGTTCACCGTGTGATCGTGAACCGGTGCGCCAG	
gp160.modSF162.delV2	(1970)	GCCTGGGCATCGTGTTCACCGTGTGATCGTGAACCGGTGCGCCAG	
gp160.modSF162.delV1V2	(1862)	GCCTGGGCATCGTGTTCACCGTGTGATCGTGAACCGGTGCGCCAG	
gp140.modSF162	(2026)		
gp140.mut.modSF162	(2026)		
gp140.mut7.modSF162	(2026)		
gp140.mut8.modSF162	(2026)		
gp120.modSF162	(1513)		
Consensus	(2051)		
gp160.modSF162	(2101)	GGCTACAGCCCTGAGCTTCCAGACCCGCTTCCCCGCCCCCGCGGCC	2150
gp160.modSF162.delV2	(2020)	GGCTACAGCCCTGAGCTTCCAGACCCGCTTCCCCGCCCCCGCGGCC	
gp160.modSF162.delV1V2	(1912)	GGCTACAGCCCTGAGCTTCCAGACCCGCTTCCCCGCCCCCGCGGCC	
gp140.modSF162	(2026)		
gp140.mut.modSF162	(2026)		
gp140.mut7.modSF162	(2026)		
gp140.mut8.modSF162	(2026)		
gp120.modSF162	(1513)		
Consensus	(2101)		
gp160.modSF162	(2151)	CGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGGACCGCGACC	2200
gp160.modSF162.delV2	(2070)	CGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGGACCGCGACC	
gp160.modSF162.delV1V2	(1962)	CGACCGCCCCGAGGGCATCGAGGAGGAGGGCGGCGGACCGCGACC	
gp140.modSF162	(2026)		
gp140.mut.modSF162	(2026)		
gp140.mut7.modSF162	(2026)		
gp140.mut8.modSF162	(2026)		
gp120.modSF162	(1513)		
Consensus	(2151)		

FIG. 66A-15

SUBSTITUTE SHEET (RULE 26)

94 / 131

	2201		2250
gp160.modSF162	(2201)	GCAGCAGCCCCCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTG	
gp160.modSF162.delV2	(2120)	GCAGCAGCCCCCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTG	
gp160.modSF162.delV1V2	(2012)	GCAGCAGCCCCCTGGTGCACGGCCTGCTGGCCCTGATCTGGGACGACCTG	
gp140.modSF162	(2026)	-----	
gp140.modSF162	(2026)	-----	
gp140.mut.modSF162	(2026)	-----	
gp140.mut7.modSF162	(2026)	-----	
gp140.mut8.modSF162	(2026)	-----	
gp120.modSF162	(1513)	-----	
Consensus	(2201)	-----	
	2251		2300
gp160.modSF162	(2251)	CGCAGCCTGTGCCTGTTTCAGCTACCAACCGCCTGCGCGACCTGATCCTGAT	
gp160.modSF162.delV2	(2170)	CGCAGCCTGTGCCTGTTTCAGCTACCAACCGCCTGCGCGACCTGATCCTGAT	
gp160.modSF162.delV1V2	(2062)	CGCAGCCTGTGCCTGTTTCAGCTACCAACCGCCTGCGCGACCTGATCCTGAT	
gp140.modSF162	(2026)	-----	
gp140.mut.modSF162	(2026)	-----	
gp140.mut7.modSF162	(2026)	-----	
gp140.mut8.modSF162	(2026)	-----	
gp120.modSF162	(1513)	-----	
Consensus	(2251)	-----	
	2301		2350
gp160.modSF162	(2301)	CGCCGCCCCGATCGTGGAGCTGCTGGGCGCCCGGGCTGGGAGGCCCTGA	
gp160.modSF162.delV2	(2220)	CGCCGCCCCGATCGTGGAGCTGCTGGGCGCCCGGGCTGGGAGGCCCTGA	
gp160.modSF162.delV1V2	(2112)	CGCCGCCCCGATCGTGGAGCTGCTGGGCGCCCGGGCTGGGAGGCCCTGA	

FIG. 66A-16

SUBSTITUTE SHEET (RULE 26)



95 / 131

gp140.modSF162	(2026)	-----	2351	2400
gp140.mut.modSF162	(2026)	-----	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC	
gp140.mut7.modSF162	(2026)	-----	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC	
gp140.mut8.modSF162	(2026)	-----	AGTACTGGGGCAACCTGCTGCAGTACTGGATCCAGGAGCTGAAGAACAGC	
gp120.modSF162	(1513)	-----	-----	
Consensus	(2301)	-----	-----	
gp160.modSF162	(2351)	-----	-----	
gp160.modSF162.delV2	(2270)	-----	-----	
gp160.modSF162.delV1V2	(2162)	-----	-----	
gp140.modSF162	(2026)	-----	-----	
gp140.mut.modSF162	(2026)	-----	-----	
gp140.mut7.modSF162	(2026)	-----	-----	
gp140.mut8.modSF162	(2026)	-----	-----	
gp120.modSF162	(1513)	-----	-----	
Consensus	(2351)	-----	-----	
gp160.modSF162	(2401)	-----	2401	2450
gp160.modSF162.delV2	(2320)	-----	GCCGTGAGCCTGTTTCGACGCCATCGCCATCGCCGTGGCCGAGGGCACCGA	
gp160.modSF162.delV1V2	(2212)	-----	GCCGTGAGCCTGTTTCGACGCCATCGCCATCGCCGTGGCCGAGGGCACCGA	
gp140.modSF162	(2026)	-----	GCCGTGAGCCTGTTTCGACGCCATCGCCATCGCCGTGGCCGAGGGCACCGA	
gp140.mut.modSF162	(2026)	-----	-----	
gp140.mut7.modSF162	(2026)	-----	-----	
gp140.mut8.modSF162	(2026)	-----	-----	
gp120.modSF162	(1513)	-----	-----	
Consensus	(2401)	-----	-----	

FIG. 66A-17

SUBSTITUTE SHEET (RULE 26)

96 / 131

gp160.modSF162	(2451)	2451	2500
gp160.modSF162.delV2	(2370)		
gp160.modSF162.delV1V2	(2262)		
gp140.modSF162	(2026)		
gp140.mut.modSF162	(2026)		
gp140.mut7.modSF162	(2026)		
gp140.mut8.modSF162	(2026)		
gp120.modSF162	(1513)		
Consensus	(2451)		
gp160.modSF162	(2501)	2501	2547
gp160.modSF162.delV2	(2420)		
gp160.modSF162.delV1V2	(2312)		
gp140.modSF162	(2026)		
gp140.mut.modSF162	(2026)		
gp140.mut7.modSF162	(2026)		
gp140.mut8.modSF162	(2026)		
gp120.modSF162	(1513)		
Consensus	(2501)		

FIG. 66A-18

SUBSTITUTE SHEET (RULE 26)

97 / 131

			Start of tPA	
	1		↓	40
gp160	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp160 del V1	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp160 del V2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp160 del V1-2	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp 160 del 128-194	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp140TM	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp140	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp140mut	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
gp120	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		
Consensus	(1)	GAATTCGCCACCATGGATGCAATGAAGAGAGGGGCTCTGCT		80
	41			
gp160	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp160 del V1	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp160 del V2	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp160 del V1-2	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp 160 del 128-194	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp140TM	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp140	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp140mut	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
gp120	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
Consensus	(41)	GTGTGCTGCTGCTGTGTGGAGCAGTCTTCGTTTCGCCCCAG		
	end of tPA		↓	120
	81			
gp160	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp160 del V1	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp160 del V2	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp160 del V1-2	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp 160 del 128-194	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp140TM	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp140	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp140mut	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
gp120	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		
Consensus	(81)	CGCCACCACCGTGCTGTGGGTGACCGTGTAACGGCGTG		160
	121			
gp 160	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp160 del V1	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp160 del V2	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp160 del V1-2	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp 160 del 128-194	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp140TM	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp140	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp140mut	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
gp120	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		
Consensus	(121)	CCCGTGTGGAAGGAGGCCACCACCACCCTGTTCTGCGCCA		

FIG. 66B-1

SUBSTITUTE SHEET (RULE 26)

		98 / 131	200
		161	
gp160	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp160 del V1	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp160 del V2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp160 del V1-2	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp 160 del 128-194	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp140TM	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp140	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp140mut	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
gp120	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
Consensus	(161)	GCGACGCCAAGGCTTACAAGGCCGAGGCCCAACAACGTGTG	
		201	240
gp160	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V1	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V2	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp160 del V1-2	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp 160 del 128-194	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140TM	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp140mut	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
gp120	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
Consensus	(201)	GGCCACCCACGCCTGCGTGCCACCGACCCCAACCCCCAG	
		241	280
gp160	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp160 del V1	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp160 del V2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp160 del V1-2	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp 160 del 128-194	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp140TM	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp140	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp140mut	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
gp120	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
Consensus	(241)	GAGGTGAACCTGACCAACGTGACCGAGAACTTCAACATGT	
		281	320
gp160	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V1	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V2	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp160 del V1-2	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp 160 del 128-194	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140TM	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp140mut	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
gp120	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
Consensus	(281)	GGAAGAACAACATGGTGGAGCAGATGCATGAGGACATCAT	
		321	360
gp160	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V2	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp160 del V1-2	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp 160 del 128-194	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140TM	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp140mut	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
gp120	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	
Consensus	(321)	CAGCCTGTGGGACCAGAGCCTGAAGCCCTGCGTGAAGCTG	

FIG. 66B-2

SUBSTITUTE SHEET (RULE 26)

	99/131	400
	361	
gp160	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
gp160 del V1	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGG	
gp160 del V2	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
gp160 del V1-2	(361) GGC-----	
gp 160 del 128-194	(361) ACCCCCCTGTGCGTGCGGGGCGAGGG-----	
gp140TM	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
gp140	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
gp140mut	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
gp120	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	
Consensus	(361) ACCCCCCTGTGCGTGACCCTGAACTGCACCGACAAGCTGA	440
	401	
gp160	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
gp160 del V1	(401) GCGCCGGC-----	
gp160 del V2	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
gp140	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
gp140mut	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
gp120	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	
Consensus	(401) CCGGCAGCACCAACGGCACCAACAGCACCAGCGGCACCAA	480
	441	
gp160	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
gp160 del V1	(409) -----	
gp160 del V2	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
gp140	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
gp140mut	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
gp120	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	
Consensus	(441) CAGCACCAGCGGCACCAACAGCACCAGCACCAACAGCACC	520
	481	
gp160	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
gp160 del V1	(409) -----GGCGAGATCAAGAACT	
gp160 del V2	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
gp160 del V1-2	(364) -----	
gp 160 del 128-194	(385) -----	
gp140TM	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
gp140	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
gp140mut	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
gp120	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	
Consensus	(481) GACAGCTGGGAGAAGATGCCCCGAGGGCGAGATCAAGAACT	560
	521	
gp160	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
gp160 del V1	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
gp160 del V2	(521) GCAGCTTCAACATCGGCGCCGCG-----	
gp160 del V1-2	(521) -----	
gp 160 del 128-194	(521) -----	
gp140TM	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
gp140	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
gp140mut	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
gp120	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	
Consensus	(521) GCAGCTTCAACATCACCACCAGCGTGCGCGACAAGGTGCA	

FIG. 66B-3

SUBSTITUTE SHEET (RULE 26)

100 / 131

		561		600
gp160	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
gp160 del V1	(465)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
gp160 del V2	(544)	-----		
gp160 del V1-2	(364)	-----		
gp 160 del 128-194	(385)	-----		
gp140TM	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
gp140	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
gp140mut	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
gp120	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
Consensus	(561)	GAAGGAGTACAGCCTGTTCTACAAGCTGGACGTGGTGCCC		
		601		640
gp160	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
gp160 del V1	(505)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
gp160 del V2	(544)	-----CGCCTGATCAACTGCA		
gp160 del V1-2	(364)	-----		
gp 160 del 128-194	(385)	-----AACTGCG		
gp140TM	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
gp140	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
gp140mut	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
gp120	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
Consensus	(601)	ATCGACAACGACAACGCCAGCTACCGCCTGATCAACTGCA		
		641		680
gp160	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp160 del V1	(545)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp160 del V2	(560)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp160 del V1-2	(364)	-----CAGGCCTGCCCCAAGGTGAGCTT		
gp 160 del 128-194	(392)	AGACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp140TM	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp140	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp140mut	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
gp120	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
Consensus	(641)	ACACCAGCGTGATCACCAGGCCTGCCCCAAGGTGAGCTT		
		681		720
gp160	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp160 del V1	(585)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp160 del V2	(600)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp160 del V1-2	(387)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp 160 del 128-194	(432)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp140TM	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp140	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp140mut	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
gp120	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
Consensus	(681)	CGAGCCCATCCCCATCCACTACTGCGCCCCCGCCGGCTTC		
		721		760
gp160	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp160 del V1	(625)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp160 del V2	(640)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp160 del V1-2	(427)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp 160 del 128-194	(472)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp140TM	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp140	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp140mut	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
gp120	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		
Consensus	(721)	GCCATCCTGAAGTGCAAGGACAAGAAGTTCAACGGCACCG		

FIG. 66B-4

SUBSTITUTE SHEET (RULE 26)

101 / 131

		761	800
gp160	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp160 del V1	(665)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp160 del V2	(680)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp160 del V1-2	(467)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp 160 del 128-194	(512)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp140TM	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp140	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp140mut	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
gp120	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	
Consensus	(761)	GCCCCTGCAAGAACGTGAGCACCGTGAGTGCACCCACGG	840
		801	
gp160	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1	(705)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V2	(720)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp160 del V1-2	(507)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp 160 del 128-194	(552)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140TM	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp140mut	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
gp120	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	
Consensus	(801)	CATCCGCCCCGTGGTGAGCACCCAGCTGCTGCTGAACGGC	880
		841	
gp160	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp160 del V1	(745)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp160 del V2	(760)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp160 del V1-2	(547)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp 160 del 128-194	(592)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp140TM	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp140	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp140mut	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
gp120	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	
Consensus	(841)	AGCCTGGCCGAGGAGGAGATCGTGCTGCGCTCCGAGAACT	920
		881	
gp160	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp160 del V1	(785)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp160 del V2	(800)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp160 del V1-2	(587)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp 160 del 128-194	(632)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp140TM	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp140	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp140mut	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
gp120	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	
Consensus	(881)	TCACCGACAACGCCAAGACCATCATCGTGAGCTGAACGA	960
		921	
gp160	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp160 del V1	(825)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp160 del V2	(840)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp160 del V1-2	(627)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp 160 del 128-194	(672)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp140TM	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp140	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp140mut	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
gp120	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	
Consensus	(921)	GTCCGTGGAGATCAACTGCATCCGCCCCAACAACAACACG	

FIG. 66B-5

SUBSTITUTE SHEET (RULE 26)

102 / 131

	961	1000
gp160	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp160 del V1	(865) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp160 del V2	(880) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp160 del V1-2	(667) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp 160 del 128-194	(712) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp140TM	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp140	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp140mut	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
gp120	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
Consensus	(961) CGTAAGAGCATCCACATCGGCCCCGGCCGCGCCTTCTACG	
	1001	1040
gp160	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V1	(905) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V2	(920) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp160 del V1-2	(707) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp 160 del 128-194	(752) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140TM	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp140mut	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
gp120	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
Consensus	(1001) CCACCGGCGACATCATCGGCGACATCCGCCAGGCCCACTG	
	1041	1080
gp160	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V1	(945) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V2	(960) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp160 del V1-2	(747) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp 160 del 128-194	(792) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140TM	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp140mut	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
gp120	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
Consensus	(1041) CAACATCAGCAAGGCCAACTGGACCAACACCCTCGAGCAG	
	1081	1120
gp160	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp160 del V1	(985) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp160 del V2	(1000) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp160 del V1-2	(787) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp 160 del 128-194	(832) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp140TM	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp140	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp140mut	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
gp120	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
Consensus	(1081) ATCGTGGAGAAGCTGCGCGAGCAGTTCGGCAACAACAAGA	
	1121	1160
gp160	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V1	(1025) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V2	(1040) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp160 del V1-2	(827) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp 160 del 128-194	(872) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140TM	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp140mut	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
gp120	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	
Consensus	(1121) CCATCATCTTCAACAGCAGCAGCGGCGGCGACCCCGAGAT	

FIG. 66B-6

SUBSTITUTE SHEET (RULE 26)



103 / 131

		1161		1200
gp160	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp160 del V1	(1065)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp160 del V2	(1080)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp160 del V1-2	(867)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp 160 del 128-194	(912)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp140TM	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp140	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp140mut	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
gp120	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
Consensus	(1161)	CGTGTTCACAGCTTCAACTGCGGCGGCGAGTTCTTCTAC		
		1201		1240
gp160	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp160 del V1	(1105)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp160 del V2	(1120)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp160 del V1-2	(907)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp 160 del 128-194	(952)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp140TM	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp140	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp140mut	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
gp120	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
Consensus	(1201)	TGCAACACCAGCCAGCTGTTCAACAGCACCTGGAACATCA		
		1241		1280
gp160	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V1	(1145)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V2	(1160)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp160 del V1-2	(947)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp 160 del 128-194	(992)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140TM	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp140mut	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
gp120	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
Consensus	(1241)	CCGAGGAGGTGAACAAGACCAAGGAGAACGACACCATCAT		
		1281		1320
gp160	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V1	(1185)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V2	(1200)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp160 del V1-2	(987)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp 160 del 128-194	(1032)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp140TM	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp140	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp140mut	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
gp120	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
Consensus	(1281)	CCTGCCCTGCCGCATCCGCCAGATCATCAACATGTGGCAG		
		1321		1360
gp160	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp160 del V1	(1225)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp160 del V2	(1240)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp160 del V1-2	(1027)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp 160 del 128-194	(1072)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp140TM	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp140	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp140mut	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
gp120	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		
Consensus	(1321)	GAGGTGGGCAAGGCCATGTACGCCCCCCCCATCCGCGGCC		

FIG. 66B-7

SUBSTITUTE SHEET (RULE 26)

104 / 131

		1361	1400
gp160	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V1	(1265)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V2	(1280)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp160 del V1-2	(1067)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp 160 del 128-194	(1112)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140TM	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp140mut	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
gp120	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
Consensus	(1361)	AGATCAAGTGCAGCAGCAATATTACCGGCCTGCTGCTGAC	
		1401	1440
gp160	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp160 del V1	(1305)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp160 del V2	(1320)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp160 del V1-2	(1107)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp 160 del 128-194	(1152)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp140TM	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp140	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp140mut	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
gp120	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
Consensus	(1401)	CCGCGACGGCGGCACCAACAACAACCGCACCAACGACACC	
		1441	1480
gp160	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp160 del V1	(1345)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp160 del V2	(1360)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp160 del V1-2	(1147)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp 160 del 128-194	(1192)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp140TM	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp140	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp140mut	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
gp120	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
Consensus	(1441)	GAGACCTTCCGCCCCGGCGGGCGGCAACATGAAGGACAAC	
		1481	1520
gp160	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V1	(1385)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V2	(1400)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp160 del V1-2	(1187)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp 160 del 128-194	(1232)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140TM	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp140mut	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
gp120	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
Consensus	(1481)	GGCGCAGCGAGCTGTACAAGTACAAGGTGGTGCGCATCGA	
		1521	1560
gp160	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp160 del V1	(1425)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp160 del V2	(1440)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp160 del V1-2	(1227)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp 160 del 128-194	(1272)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp140TM	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp140	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp140mut	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
gp120	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	
Consensus	(1521)	GCCCCCTGGGCGTGGCCCCCACCAGGCCAAGCGCCGCGTG	

FIG. 66B-8

SUBSTITUTE SHEET (RULE 26)

105 / 131

		1561		1600
			GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160	(1561)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V1	(1465)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V2	(1480)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp160 del V1-2	(1267)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp 160 del 128-194	(1312)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140TM	(1561)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140	(1561)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
gp140mut	(1561)		GTGCAGCGCGAGAAGAGCGCCGTGGGCCTGGGCGCCCTGT	
gp120	(1561)		GTGCAGCGCGAGAAGCGCTAAG-----	
Consensus	(1561)		GTGCAGCGCGAGAAGCGCGCCGTGGGCCTGGGCGCCCTGT	
		1601		1640
gp160	(1601)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V1	(1505)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V2	(1520)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp160 del V1-2	(1307)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp 160 del 128-194	(1352)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140TM	(1601)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140	(1601)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp140mut	(1601)		TCATCGGCTTC-CTGGGCGCCGCCGGGAGCACCATGGGCG	
gp120	(1583)		ATATCGGATCCTCTAGA-----	
Consensus	(1601)		TCATCGGCTTCNCTGGGCGCCGCCGGGAGCACCATGGGCG	
		1641		1680
gp160	(1640)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V1	(1544)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V2	(1559)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp160 del V1-2	(1346)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp 160 del 128-194	(1391)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140TM	(1640)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140	(1640)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp140mut	(1640)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
gp120	(1600)		-----	
Consensus	(1641)		CCGCCTCCGTGACCCTGACCGTGCAGGCCCGCCAGCTGCT	
		1681		1720
gp160	(1680)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp160 del V1	(1584)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp160 del V2	(1599)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp160 del V1-2	(1386)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp 160 del 128-194	(1431)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp140TM	(1680)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp140	(1680)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp140mut	(1680)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
gp120	(1600)		-----	
Consensus	(1681)		GAGCGGCATCGTGCAGCAGCAGAACAACCTGCTGCGCGCC	
		1721		1760
gp160	(1720)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1	(1624)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V2	(1639)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp160 del V1-2	(1426)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp 160 del 128-194	(1471)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140TM	(1720)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140	(1720)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp140mut	(1720)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	
gp120	(1600)		-----	
Consensus	(1721)		ATCGAGGCCCAGCAGCACCTGCTGCAGCTGACCGTGTGGG	

FIG. 66B-9

SUBSTITUTE SHEET (RULE 26)

106/131

		1761	1800
gp160	(1760)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp160 del V1	(1664)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp160 del V2	(1679)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp160 del V1-2	(1466)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp 160 del 128-194	(1511)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp140TM	(1760)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp140	(1760)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp140mut	(1760)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
gp120	(1600)	-----	
Consensus	(1761)	GCATCAAGCAGCTGCAGGCCCCGCATCCTGGCCGTGGAGCG	
		1801	1840
gp160	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1	(1704)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V2	(1719)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp160 del V1-2	(1506)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp 160 del 128-194	(1551)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140TM	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp140mut	(1800)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
gp120	(1600)	-----	
Consensus	(1801)	CTACCTGAAGGACCAGCAGCTGCTGGGCATCTGGGGCTGC	
		1841	1880
gp160	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V1	(1744)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V2	(1759)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp160 del V1-2	(1546)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp 160 del 128-194	(1591)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140TM	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp140mut	(1840)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
gp120	(1600)	-----	
Consensus	(1841)	AGCGGCAAGCTGATCTGCACCACCACCGTGCCCTGGAACA	
		1881	1920
gp160	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1	(1784)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V2	(1799)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp160 del V1-2	(1586)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp 160 del 128-194	(1631)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140TM	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp140mut	(1880)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
gp120	(1600)	-----	
Consensus	(1881)	GCAGCTGGAGCAACAAGAGCCTGACCGAGATCTGGGACAA	
		1921	1960
gp160	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1	(1824)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V2	(1839)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp160 del V1-2	(1626)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp 160 del 128-194	(1671)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140TM	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp140mut	(1920)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	
gp120	(1600)	-----	
Consensus	(1921)	CATGACCTGGATGGAGTGGGAGCGCGAGATCGGCAACTAC	

FIG. 66B-10

SUBSTITUTE SHEET (RULE 26)

107 / 131

		1961	2000
gp160	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp160 del V1	(1864)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp160 del V2	(1879)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp160 del V1-2	(1666)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp 160 del 128-194	(1711)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp140TM	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp140	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp140mut	(1960)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
gp120	(1600)	-----	
Consensus	(1961)	ACCGGCCTGATCTACAACCTGATCGAGATCGCCCAGAACC	
		2001	2040
gp160	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1	(1904)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V2	(1919)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp160 del V1-2	(1706)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp 160 del 128-194	(1751)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140TM	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp140mut	(2000)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
gp120	(1600)	-----	
Consensus	(2001)	AGCAGGAGAAGAACGAGCAGGAGCTGCTGGAGCTGGACAA	
		2041	2080
gp160	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp160 del V1	(1944)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp160 del V2	(1959)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp160 del V1-2	(1746)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp 160 del 128-194	(1791)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp140TM	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp140	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp140mut	(2040)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
gp120	(1600)	-----	
Consensus	(2041)	GTGGGCCAGCCTGTGGAACCTGGTTTCGACATCACCAACTGG	
		2081	2120
gp160	(2080)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V1	(1984)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V2	(1999)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp160 del V1-2	(1786)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp 160 del 128-194	(1831)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp140TM	(2080)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
gp140	(2080)	CTGTGGTACATC-----	
gp140mut	(2080)	CTGTGGTACATC-----	
gp120	(1600)	-----	
Consensus	(2081)	CTGTGGTACATCCGCATCTTCATCATGATCGTGGGCGGCC	
		2121	2160
gp160	(2120)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V1	(2024)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V2	(2039)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp160 del V1-2	(1826)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp 160 del 128-194	(1871)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCA----	
gp140TM	(2120)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCATCGT	
gp140	(2092)	-----	
gp140mut	(2092)	-----	
gp120	(1600)	-----	
Consensus	(2121)	TGATCGGCCTGCGCATCGTGTTCGCCGTGCTGAGCANNNN	

FIG. 66B-11

SUBSTITUTE SHEET (RULE 26)

108 / 131

		2161	2200
gp160	(2156)	-TCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1	(2060)	-TCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
gp160 del V2	(2075)	-TCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
gp160 del V1-2	(1862)	-TCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
gp 160 del 128-194	(1907)	-TCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
gp140TM	(2160)	GTAAGATATCGGATCCTCTAGA-----	
gp140	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp140mut	(2092)	-TAAGATATCGGATCCTCTAGA-----	
gp120	(1600)	-----	
Consensus	(2161)	NTCGTGAACCGCGTGC GCCAGGGCTACAGCCCCATCAGCC	
		2201	2240
gp160	(2195)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
gp160 del V1	(2099)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
gp160 del V2	(2114)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
gp160 del V1-2	(1901)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
gp 160 del 128-194	(1946)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2201)	TGCAGACCCGCGCTGCCCCGCCAGCGCGGCCCCGACCGCCC	
		2241	2280
gp160	(2235)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
gp160 del V1	(2139)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
gp160 del V2	(2154)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
gp160 del V1-2	(1941)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
gp 160 del 128-194	(1986)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2241)	CGAGGGCATCGAGGAGGAGGGCGGCAGCGCGACCGCGAC	
		2281	2320
gp160	(2275)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
gp160 del V1	(2179)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
gp160 del V2	(2194)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
gp160 del V1-2	(1981)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
gp 160 del 128-194	(2026)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2281)	CGCAGCAACCGCCTGGTGCACGGCCTGCTGGCCCTGATCT	
		2321	2360
gp160	(2315)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	
gp160 del V1	(2219)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	
gp160 del V2	(2234)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	
gp160 del V1-2	(2021)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	
gp 160 del 128-194	(2066)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2321)	GGGACGACCTGCGCAGCCTGTGCCTGTTTCAGCTACCACCG	

FIG. 66B-12

SUBSTITUTE SHEET (RULE 26)

109/131

2361

2400

gp160 (2355) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

gp160 del V1 (2259) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

gp160 del V2 (2274) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

gp160 del V1-2 (2061) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

gp 160 del 128-194 (2106) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

gp140TM (2182) -----

gp140 (2113) -----

gp140mut (2113) -----

gp120 (1600) -----

Consensus (2361) CCTGCGCGACCTGCTGCTGATCGTGGCCCGCATCGTGGAG

2401

2440

gp160 (2395) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

gp160 del V1 (2299) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

gp160 del V2 (2314) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

gp160 del V1-2 (2101) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

gp 160 del 128-194 (2146) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

gp140TM (2182) -----

gp140 (2113) -----

gp140mut (2113) -----

gp120 (1600) -----

Consensus (2401) CTGCTGGGCGCGCCGCGGCTGGGAGGCCCTGAAGTACTGGT

2480

2441

gp160 (2435) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

gp160 del V1 (2339) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

gp160 del V2 (2354) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

gp160 del V1-2 (2141) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

gp 160 del 128-194 (2186) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

gp140TM (2182) -----

gp140 (2113) -----

gp140mut (2113) -----

gp120 (1600) -----

Consensus (2441) GGAACCTGCTGCAGTACTGGAGCCAGGAGCTGAAGAGCAG

2520

2481

gp160 (2475) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

gp160 del V1 (2379) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

gp160 del V2 (2394) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

gp160 del V1-2 (2181) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

gp 160 del 128-194 (2226) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

gp140TM (2182) -----

gp140 (2113) -----

gp140mut (2113) -----

gp120 (1600) -----

Consensus (2481) CGCCGTGAGCCTGTTCAACGCCACCGCCATCGCCGTGGCC

2560

2521

gp160 (2515) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

gp160 del V1 (2419) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

gp160 del V2 (2434) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

gp160 del V1-2 (2221) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

gp 160 del 128-194 (2266) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

gp140TM (2182) -----

gp140 (2113) -----

gp140mut (2113) -----

gp120 (1600) -----

Consensus (2521) GAGGGCACCGACCGCATCATCGAGATCGTGCAGCGCATCT

FIG. 66B-13

SUBSTITUTE SHEET (RULE 26)

		2561	2600
gp160	(2555)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
gp160 del V1	(2459)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
gp160 del V2	(2474)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
gp160 del V1-2	(2261)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
gp 160 del 128-194	(2306)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2561)	TCCGCGCCGTGATCCACATCCCCGCCGCATCCGCCAGGG	
		2601	2640
gp160	(2595)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
gp160 del V1	(2499)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
gp160 del V2	(2514)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
gp160 del V1-2	(2301)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
gp 160 del 128-194	(2346)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2601)	CCTGGAGCGCGCCCTGCTGTAAGATATCGGATCCTCTAGA	
		2641	2680
gp160	(2635)	AAGCCATGGATATCGGATCCA CTACGCGTTAGAGCTCGCT	
gp160 del V1	(2539)	-----	
gp160 del V2	(2554)	-----	
gp160 del V1-2	(2341)	-----	
gp 160 del 128-194	(2386)	-----	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2641)	NNN	
		2681	
gp160	(2675)	GATCAGCT	
gp160 del V1	(2539)	-----	
gp160 del V2	(2554)	-----	
gp160 del V1-2	(2341)	-----	
gp 160 del 128-194	(2386)	-----	
gp140TM	(2182)	-----	
gp140	(2113)	-----	
gp140mut	(2113)	-----	
gp120	(1600)	-----	
Consensus	(2681)	NNNNNNNN	

FIG. 66B-14

***SUBSTITUTE SHEET (RULE 26)***



111 / 131

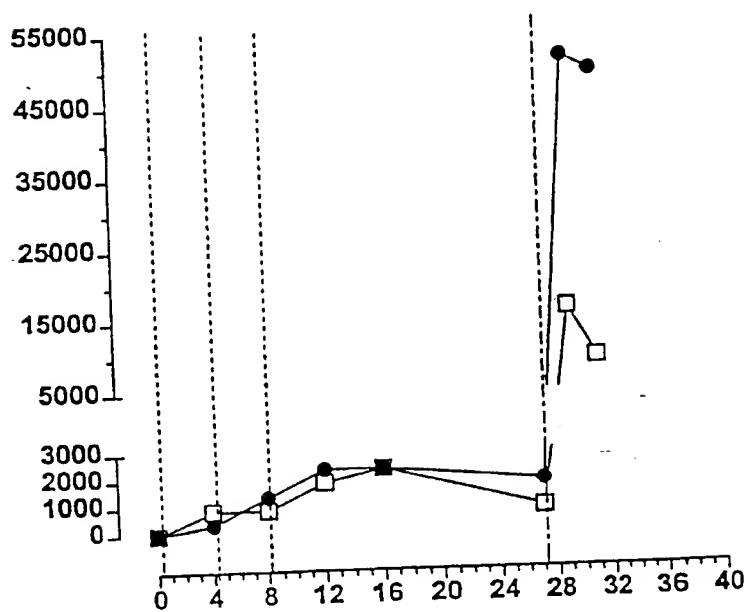


FIG. 67

SUBSTITUTE SHEET (RULE 26)

112 / 131

HIV-1SF2 wt RT (PISPIET--&gt;GIRKVL)

CCCATTAGTCCTATTGAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCCAAAA  
GTTAAGCAATGGCCATTGACAGAAGAAAAAATAAAGCATTAGTAGAGATATGTACAGAA  
ATGGAAAAGGAAGGGAAAATTTCAAAAATTGGGCCTGAAAATCCATACAATACTCCAGTA  
TTTGCTATAAAGAAAAAAGACAGTACTAAATGGAGAAAAGTAGTAGATTTTCAGAGAACTT  
AATAAAAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGAATACCACACCCCGCAGGGTTA  
AAAAAGAAAAAATCAGTAACAGTATTGGATGTGGGTGATGCATACTTTTCAGTTCCCTTA  
GATAAAGACTTTAGAAAGTATACTGCATTTACCATACTAGTATAAACAATGAGACACCA  
GGGATTAGATATCAGTACAATGTGCTGCCACAGGGATGGAAAGGATCACCAGCAATATTC  
CAAAGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAGAATCCAGACATAGTTATC  
TATCAAtacatggatgatTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACA  
AAAATAGAGGAAGTGAACAGCATCTGTTGAGGTGGGGATTTACCACACCAGACAAAAAA  
CATCAGAAAGAACCTCCATTCCTTtggatgggttatGAACTCCATCCTGATAAATGGACA  
GTACAGCCTATAATGCTGCCAGAAAAAGACAGCTGGACTGTCAATGACATACAGAAGTTA  
GTGGGAAAATTGAATTGGGCAAGTCAGATTTATGCAGGGATTAAAGTAAAGCAGTTATGT  
AAACTCCTTAGAGGAACCAAAGCACTAACAGAAGTAATACCACTAACAGAAGAAGCAGAG  
CTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAACCAGTACATGAAGTATATTATGAC  
CCATCAAAAGACTTAGTAGCAGAAATACAGAAGCAGGGGCAAGGQCAATGGACATATCAA  
ATTTATCAAGAGCCATTTAAAAATCTGAAAACAGGAAAGTATGCAAGGATGAGGGGTGCC  
CACACTAATGATGTAAACAGTTAACAGAGGCAGTGCAAAAAGTATCCACAGAAAGCATA  
GTAATATGGGGAAAGATTCCATAAATTTAAACTACCCATACAAAAGGAAACATGGGAAGCA  
TGGTGGATGGAGTATTGGCAAGCTACCTGGATTCCCTGAGTGGGAGTTTGTCAATACCCCT  
CCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAACCCATAGTAGGAGCAGAACTTTC  
TATGTAGATGGGGCAGCTAATAGGGAGACTAAATTAGGAAAAGCAGGATATGTTACTGAC  
AGAGGAAGACAAAAAGTTGTCTCCATAGCTGACACAACAAATCAGAAGACTGAATTACAA  
GCAATTCATCTAGCTTTGCAGGATTGGGGATTAGAAGTAAACATAGTAACAGACTCACAA  
TATGCATTAGGAATCATTCAAGCACAAACCAGATAAGAGTGAATCAGAGTTAGTCAGTCAA  
ATAATAGAGCAGTTAATAAAAAAGGAAAAGGTCTACCTGGCATGGGTACCAGCACACAAA  
GGAATTGGAGGAAATGAACAAGTAGATAAATTAGTCAGTGCTGGAATCAGGAAAGTACTA

**FIG. 68**

(SEQ ID NO:77)

SUBSTITUTE SHEET (RULE 26)

113 / 131

GagProtMod.SF2 (GP1)

GagProtMod.SF2 (GPL)

GTTCGACGCCACCATGGGCGCCCGCGCCAGCGTGCTGAGCGGGCGGCGAGCTGGACAAAGTGG  
GAGAAGATCCGCCTGCGCCCCGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG  
GCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCCTGCTGGAGACCAGCGAGGGC  
TGCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGCGC  
AGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGCACCAGCGCATCGACGTCAAGGAC  
ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCACG  
CAGGCCGCGCGCCGCCCGGCACCGGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC  
GTGCAGAACCTGCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCCCTGAACGCC  
TGGGTGAAGGTGGTGGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTACGCGCC  
CTGAGCGAGGGCGCCACCCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCAC  
CAGGCCGCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCGCGGAGTGGGACCGC  
GTGCACCCCGTGCACGCCGGCCCCATCGCCCCCGGCCAGATGCGCGAGCCCCGCGGCAGC  
GACATCGCCGGCACCACCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCAACAACCCC  
CCCATCCCCGTGGGCGAGATCTACAAGCGGTGGATCATCCTGGGCCCTGAACAAGATCGTG  
CGGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCGC  
GACTACGTGGACCGCTTCTACAAGACCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAG  
AACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCCGACTGCAAGACCATCCTG  
AAGGCTCTCGGCCCCGCGGCCACCCTGGAGGAGATGATGACCGCTGCCAGGGCGTGGGC  
GGCCCCGGCCACAAGGCCCGCGTGTGGCCGAGGCGATGAGCCAGGTGACGAACCCGGCG  
ACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAAC  
TGCGGCAAGGAGGGCCACACCGCCAGGAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGG  
CGCTGCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTA  
GGGAAGATCTGGCCTTCTTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAG  
CCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACTCCCTCTCAG  
AAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGC  
AACGACCCCTCGTCACAGTAAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCG  
GCGCCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAAGCCCAAGATGA  
TCGGCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCT  
GCGGCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCC  
GCAACCTGCTGACCCAGATCGGCTGCACCTGAACTTCCCCATCAGCCCCATCGAGACGG  
TGCCCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCCTGTAAG  
AATTC

FIG. 62

FIG. 69

(SEQ ID NO:78)

***SUBSTITUTE SHEET (RULE 26)***

114 / 131

## GagProtMod.SF2 (GP2)

GTCGACGCCACCATGGGCGCCCGCGCCAGCGTGCTGAGCGGCGGCGAGCTGGACAAGTGG  
GAGAAGATCCGCCTGCGCCCCGGCGGCAAGAAGAAGTACAAGCTGAAGCACATCGTGTGG  
GCCAGCCGCGAGCTGGAGCGCTTCGCCGTGAACCCCGGCCTGCTGGAGACCAGCGAGGGC  
TGCCGCCAGATCCTGGGCCAGCTGCAGCCCAGCCTGCAGACCGGCAGCGAGGAGCTGCGC  
AGCCTGTACAACACCGTGGCCACCCTGTACTGCGTGCACCAGCGCATCGACGTCAAGGAC  
ACCAAGGAGGCCCTGGAGAAGATCGAGGAGGAGCAGAACAAGTCCAAGAAGAAGGCCAG  
CAGGCCGCGCCGCCCGCCGGCACC GGCAACAGCAGCCAGGTGAGCCAGAACTACCCCATC  
GTGCAGAACCTGCAGGGCCAGATGGTGCACCAGGCCATCAGCCCCCGCACCCCTGAACGCC  
TGGGTGAAGGTGGTGGAGGAGAAGGCCTTCAGCCCCGAGGTGATCCCCATGTTTCAGCGCC  
CTGAGCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCGTGGGCGGCCAC  
CAGGCCGCCATGCAGATGCTGAAGGAGACCATCAACGAGGAGGCCGCCGAGTGGGACCGC  
GTGCACCCCGTGCACGCCGGCCCCATCGCCCCCGGCCAGATGCGCGAGCCCCCGCGGCAGC  
GACATCGCCGGCACCACCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCAACAACCC  
CCCATCCCCGTGGGCGAGATCTACAAGCGGTGGATCATCCTGGGCCTGAACAAGATCGTG  
CGGATGTACAGCCCCACCAGCATCCTGGACATCCGCCAGGGCCCCAAGGAGCCCTTCCGC  
GACTACGTGGACCGCTTCTACAAGACCCTGCGCGCTGAGCAGGCCAGCCAGGACGTGAAG  
AACTGGATGACCGAGACCCTGCTGGTGCAGAACGCCAACCCCGACTGCAAGACCATCCTG  
AAGGCTCTCGGCCCCGCGGCCACCCTGGAGGAGATGATGACCGCTGCCAGGGCGTGGGC  
GGCCCCGGCCACAAGGCCCGCGTGCTGGCCGAGGCGATGAGGCAGGTGACGAACCCGGCG  
ACCATCATGATGCAGCGCGGCAACTTCCGCAACCAGCGGAAGACCGTCAAGTGCTTCAAC  
TGCGGCAAGGAGGGCCACACCGCCAGGAACTGCCGCGCCCCCGCAAGAAGGGCTGCTGG  
CGCTGCGGCCCGGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTA  
GGGAAGATCTGGCCTTCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAG  
CCAACAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAG  
AAGCAGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGC  
AACGACCCCTCGTCAAGTAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAG  
GAGCAGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAAATGA  
TAGGGGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCT  
GTGGACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAA  
GAAATCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGG  
TGCCCGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGTAAG  
AATTC

FIG. 70

(SEQ ID NO:79)

SUBSTITUTE SHEET (RULE 26)

115 / 131

## FS(+)\_ProtInact\_RTopt\_YM

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGGA  
AGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAA  
CAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAGC  
AGGAGCCGATAGACAAGGAACTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACG  
ACCCCTCGTCAACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC  
AGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAACCAAAAATGATAGG  
GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTGG  
ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA  
TCTGTTGACCCAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGGTGCC  
CGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGGA  
GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA  
GATCGGCCCCGAGAACCCTACAACACCCCCGTGTTCCGCATCAAGAAGAAGGACAGCAC  
CAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA  
GGTGCAGCTGGGCATCCCCACCCCCGCGGCTGAAGAAGAAGAAGAGCGTGACCGTGCT  
GGACGTGGGCGACGCCTACTTCAGCGTGCCCCCTGGACAAGGACTTCCGCAAGTACACCGC  
CTTCACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGCT  
GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGA  
GCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCTGTACGTGGGCAG  
CGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCG  
CTGGGGCTTCACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGGG  
CTACGAGCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCCGAGAAGGACAG  
CTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTA  
CGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGA  
GGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAA  
GGAGCCCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAA  
GCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGAC  
CGGCAAGTACGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGC  
CGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCT

FIG. 71A

(SEQ ID NO:80)

SUBSTITUTE SHEET (RULE 26)

116 / 131

GCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGAT  
CCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAA  
GGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACC GCGAGACCAA  
GCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCGGCAGAAAGGTGGTGAGCATCGCCGA  
CACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCT  
GGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGA  
CAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGT  
GTACCTGGCCTGGGTGCCCCGCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCT  
GGTGAGCGCCGGCATCCGCAAGGTGCTGTTCCTGAACGGCATCGATGGCGGCATCGTGAT  
CTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAGCT  
TCCCGGGGCTAGCACCGGTGAATTC

**FIG. 71B**  
(SEQ ID NO:80)

*SUBSTITUTE SHEET (RULE 26)*

117 / 131

**FS(+)\_ProtInact\_RTopt\_YMWM**

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTTTTAGGGA  
AGATCTGGCCTTCCTACAAGGGAAGGCCAGGGAATTTTCTTCAGAGCAGACCAGAGCCAA  
CAGCCCCACCAGAAGAGAGCTTCAGGTTTGGGGAGGAGAAAACAACTCCCTCTCAGAAGC  
AGGAGCCGATAGACAAGGAAGTGTATCCTTTAACTTCCCTCAGATCACTCTTTGGCAACG  
ACCCCTCGTCACAATAAGGATCGGGGGGCAACTCAAGGAAGCGCTGCTCGATACAGGAGC  
AGATGATACAGTATTAGAAGAAATGAATTTGCCAGGAAAATGGAAACCAAAATGATAGG  
GGGGATCGGGGGCTTCATCAAGGTGAGGCAGTACGACCAGATACCTGTAGAAATCTGTGG  
ACATAAAGCTATAGGTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAA  
TCTGTTGACCCAGATCGGCTGCACCTTGAAGTTCCCATCAGCCCTATTGAGACGGTGCC  
CGTGAAGTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCCATTGACCGAGGA  
GAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCAA  
GATCGGCCCCGAGAACCCCTACAACACCCCCGTGTTCCGCATCAAGAAGAAGGACAGCAC  
CAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGGA  
GGTGCAGCTGGGCATCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGCT  
GGACGTGGGCGACGCTACTTCAGCGTGCCCTGGACAAGGACTTCCGCAAGTACACCGC  
CTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGCT  
GCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGGA  
GCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCTGTACGTGGGCAG  
CGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGCG  
CTGGGGCTTCACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCCTGCCCATCGA  
GCTGCACCCCGACAAGTGGACCGTGCAGCCCATCATGCTGCCCGAGAAGGACAGCTGGAC  
CGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCGG  
CATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCGAGGTGAT  
CCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGCC  
CGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGGG  
CCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCAA  
GTACGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGCA  
GAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCAT

**FIG. 72A**

(SEQ ID NO:81)

*SUBSTITUTE SHEET (RULE 26)*

118 / 131

CCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCGA  
GTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGCC  
CATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACC GCGAGACCAAGCTGGG  
CAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCAC  
CAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGGT  
GAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCCAGCCCGACAAGAG  
CGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACCT  
GGCCTGGGTGCCCCGCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGAG  
CGCCGGCATCCGCAAGGTGCTGTTCCCTGAACGGCATCGATGGCGGCATCGTGATCTACCA  
GTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCGG  
GGCTAGCACCGGTGAATTC

**FIG. 72B**

(SEQ ID NO:81)

*SUBSTITUTE SHEET (RULE 26)*



119 / 131

## FS(-)\_ProtMod\_RTpt\_YM

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCC  
AGGACCTGGCCTTCCTGCAGGGCAAGGCCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA  
ACAGCCCCACCCGCCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG  
CCGGCGCCGACCGCCAGGGCACCCTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC  
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG  
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCCGCAAGTGGAAGCCCAAGATGATCG  
GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG  
GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCCGTGAACATCATCGGCCGCA  
ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTG  
CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG  
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA  
AGATCGGCCCCGAGAACCCTACAACACCCCCGTGTTCCGCATCAAGAAGAAGGACAGCA  
CCAAGTGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG  
AGGTGCAGCTGGGCATCCCCACCCCGCCGGCCTGAAGAAGAAGAAGAGCGTGACCGTGC  
TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCCTGGACAAGGACTTCCGCAAGTACACCG  
CCTTACCATCCCCAGCATCAACAACGAGACCCCCGGCATCCGCTACCAGTACAACGTGC  
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG  
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA  
GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC  
GCTGGGGCTTCACCACCCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGTGGATGG  
GCTACGAGCTGCACCCCGACAAGTGGACCGTGACGCCATCATGCTGCCCCGAGAAGGACA  
GCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCT  
ACGCCGGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCCTGACCG  
AGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGA  
AGGAGCCCGTGACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGA  
AGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGA  
CCGGCAAGTACGCCCGCATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGG  
CCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGC

FIG. 73A

(SEQ ID NO:82)

SUBSTITUTE SHEET (RULE 26)

120 / 131

TGCCCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGA  
TCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGA  
AGGAGCCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACC GCGAGACCA  
AGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCGGCGCAGAAGGTGGTGAGCATCGCCG  
ACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCC  
TGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCG  
ACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGG  
TGTACCTGGCCTGGGTGCCCCGCCACAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGC  
TGGTGAGCGCCGGCATCCGCAAGGTGCTGTTTCCTGAACGGCATCGATGGCGGCATCGTGA  
TCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAGC  
TTCCCCGGGGCTAGCACCGGTGAATTC

**FIG. 73B**

(SEQ ID NO:82)

*SUBSTITUTE SHEET (RULE 26)*

121 / 131

**FS(-)\_ProtMod\_RTpt\_YMWM**

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  
AGGACCTGGCCTTCTCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA  
ACAGCCCCACCCGCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG  
CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC  
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG  
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCCGCAAGTGGAAGCCCAAGATGATCG  
GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG  
GCCACAAGGCCATCGGCACCGTGCTGGTGGGGCCCCACCCCGTGAACATCATCGGCCGCA  
ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC  
CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG  
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA  
AGATCGGCCCCGAGAACCCTTACAACACCCCGTGTTTCGCCATCAAGAAGAAGGACAGCA  
CCAAGTGGCGGAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG  
AGGTGCAGCTGGGCATCCCCACCCCGCGGCCTGAAGAAGAAGAAGAGCGTGACCGTG  
TGGACGTGGGCGACGCCTACTTCAGCGTGCCCCTGGACAAGGACTTCCGCAAGTACACCG  
CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGC  
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCCAGAGCAGCATGACCAAGATCCTGG  
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGGCCCCCCTGTACGTGGGCA  
GCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACCTGCTGC  
GCTGGGGCTTCAACACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGCCCATCG  
AGCTGCACCCCGACAAGTGGACCGTGAGCCCATCATGCTGCCCCGAGAAGGACAGCTGGA  
CCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCCAGATCTACGCCG  
GCATCAAGGTGAAGCAGCTGTGCAAGCTGTGCGCGGCACCAAGGCCCTGACCGAGGTGA  
TCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGATCCTGAAGGAGC  
CCGTGCACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGATCCAGAAGCAGG  
GCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACCTGAAGACCGGCA  
AGTACGCCCCGATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGACCGAGGCCGTGC  
AGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGTTCAAGCTGCCCA

**FIG. 74A**

(SEQ ID NO:83)

*SUBSTITUTE SHEET (RULE 26)*

122 / 131

TCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCACCTGGATCCCCG  
AGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGCTGGAGAAGGAGC  
CCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACC GCGAGACCAAGCTGG  
GCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCATCGCCGACACCA  
CCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACAGCGGCCTGGAGG  
TGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCAGCCCGACAAGA  
GCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGGAGAAGGTGTACC  
TGGCCTGGGTGCCCCGCCCAAGGGCATCGGCGGCAACGAGCAGGTGGACAAGCTGGTGA  
GCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCATCGTGATCTACC  
AGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATTAAAAGCTTCCCCG  
GGGCTAGCACCGGTGAATTC

**FIG. 74B**

(SEQ ID NO:83)

*SUBSTITUTE SHEET (RULE 26)*

123 / 131

## FS(-)\_ProtMod\_RTpt(+)

GCGGCCGCGAAGGACACCAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG  
AGGACCTGGCCTTCCTGCAGGGCAAGGCCCGCGAGTTCAGCAGCGAGCAGACCCGCGCCA  
ACAGCCCCACCCGCGCGAGCTGCAGGTGTGGGGCGGCGAGAACAACAGCCTGAGCGAGG  
CCGGCGCCGACCGCCAGGGCACCGTGAGCTTCAACTTCCCCCAGATCACCTGTGGCAGC  
GCCCCCTGGTGACCATCAGGATCGGCGGCCAGCTCAAGGAGGCGCTGCTCGACACCGGCG  
CCGACGACACCGTGCTGGAGGAGATGAACCTGCCCGGCAAGTGGAAGCCCAAGATGATCG  
GCGGGATCGGGGGCTTCATCAAGGTGCGGCAGTACGACCAGATCCCCGTGGAGATCTGCG  
GCCACAAGGCCATCGGCACCGTGCTGGTGGGCCCCACCCCGTGAACATCATCGGCCGCA  
ACCTGCTGACCCAGATCGGCTGCACCCTGAACTTCCCCATCAGCCCCATCGAGACGGTGC  
CCGTGAAGCTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAGTGGCCCCTGACCGAGG  
AGAAGATCAAGGCCCTGGTGGAGATCTGCACCGAGATGGAGAAGGAGGGCAAGATCAGCA  
AGATCGGCCCCGAGAACCCCTACAACACCCCGTGTTGCCATCAAGAAGAAGGACAGCA  
CCAAGTGGCGCAAGCTGGTGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGG  
AGGTGCAGCTGGGCATCCCCACCCCGCGGCCCTGAAGAAGAAGAAGAGCGTGACCGTGC  
TGGACGTGGGCGACGCCTACTTCAGCGTGCCCTGGACAAGGACTTCCGCAAGTACACCG  
CCTTCACCATCCCCAGCATCAACAACGAGACCCCGGCATCCGCTACCAGTACAACGTGC  
TGCCCCAGGGCTGGAAGGGCAGCCCCGCCATCTTCAGAGCAGCATGACCAAGATCCTGG  
AGCCCTTCCGCAAGCAGAACCCCGACATCGTGATCTACCAGTACATGGACGACCTGTACG  
TGGGCAGCGACCTGGAGATCGGCCAGCACCGCACCAAGATCGAGGAGCTGCGCCAGCACC  
TGCTGCGCTGGGGCTTACCACCCCGACAAGAAGCACCAGAAGGAGCCCCCTTCTGT  
GGATGGGCTACGAGCTGCACCCCGACAAGTGGAACCGTGACGCCATCATGCTGCCCGAGA  
AGGACAGCTGGACCGTGAACGACATCCAGAAGCTGGTGGGCAAGCTGAACTGGGCCAGCC  
AGATCTACCGCGCATCAAGGTGAAGCAGCTGTGCAAGCTGCTGCGCGGCACCAAGGCCC  
TGACCGAGGTGATCCCCCTGACCGAGGAGGCCGAGCTGGAGCTGGCCGAGAACCGCGAGA  
TCCTGAAGGAGCCCGTGACGAGGTGTACTACGACCCAGCAAGGACCTGGTGGCCGAGA  
TCCAGAAGCAGGGCCAGGGCCAGTGGACCTACCAGATCTACCAGGAGCCCTTCAAGAACC  
TGAAGACCGGCAAGTACGCCCCGATGCGCGGCGCCACACCAACGACGTGAAGCAGCTGA  
CCGAGGCCGTGCAGAAGGTGAGCACCGAGAGCATCGTGATCTGGGGCAAGATCCCCAAGT  
TCAAGCTGCCATCCAGAAGGAGACCTGGGAGGCCTGGTGGATGGAGTACTGGCAGGCCA  
CCTGGATCCCCGAGTGGGAGTTCGTGAACACCCCCCCCCCTGGTGAAGCTGTGGTACCAGC  
TGGAGAAGGAGCCCATCGTGGGCGCCGAGACCTTCTACGTGGACGGCGCCGCCAACCGCG

**FIG. 75A**

(SEQ ID NO:84)

SUBSTITUTE SHEET (RULE 26)

124/131

AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAGAAGGTGGTGAGCA  
TCGCCGACACCACCAACCAGAAGACCGAGCTGCAGGCCATCCACCTGGCCCTGCAGGACA  
GCGGCCTGGAGGTGAACATCGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCCC  
AGCCCGACAAGAGCGAGAGCGAGCTGGTGAGCCAGATCATCGAGCAGCTGATCAAGAAGG  
AGAAGGTGTACCTGGCCTGGGTGCCCCGCCACAAGGGCATCGGCCGGCAACGAGCAGGTGG  
ACAAGCTGGTGAGCGCCGGCATCCGCAAGGTGCTGTTCTGAACGGCATCGATGGCGGCA  
TCGTGATCTACCAGTACATGGACGACCTGTACGTGGGCAGCGGCGGCCCTAGGATCGATT  
AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC

**FIG. 75B**

(SEQ ID NO:84)

*SUBSTITUTE SHEET (RULE 26)*

125 / 131

Tat\_wt\_SF162 (wildtype)

ATGGAGCCAGTAGATCCTAGATTAGAGCCCTGGAAGCATCCAGGAAGTCAGCCTAAGA  
CTGCTTGTAACAAATTGCTATTGTAAAAAGTGTTGCTTTTCATTGCCAAGTTTGTTCATAAC  
AAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGAGCTCCT  
CCAGACAGTGAGGTTTCATCAAGTTTCTCTACCAAAGCAACCCGCTTCCCAGCCCCAAGG  
GGACCCGACAGGCCCGAAGGAATCGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGA  
TCCAGTCCATTAG

**FIG. 76**

(SEQ ID NO:85)

Tat\_SF162

MEPVDPRLPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKGLGISYGRKKRRQRRRAPDSE  
VHQVSLPKQPASQPQGDPTGPKESKKKVERETETDPVH

**FIG. 77**

(SEQ ID NO:86)

Tat\_SF162\_opt

ATGGAGCCCGTGGACCCCCGCCTGGAGCCCTGGAAGCACCCCGGCAGCCAGCCCAAGAC  
CGCCTGCACCAACTGCTACTGCAAGAAGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACC  
AAGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCCAGCGCCGCCGCGCCCCCCCC  
CGACAGCGAGGTGCACCAGGTGAGCCTGCCCAAGCAGCCCGCCAGCCAGCCCCAGGGCG  
ACCCACCGGCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCGAGACCGACCCCG  
GTGCACTAG

**FIG. 78**

(SEQ ID NO:87)

Tat\_Cys22\_SF162\_opt

ATGGAGCCCGTGGACCCCCGCCTGGAGCCCTGGAAGCACCCCGGCAGCCAGCCCAAGAC  
CGCCGACCAACTGCTACTGCAAGAAGTGCTGCTTCCACTGCCAGGTGTGCTTCATCACCA  
AGGGCCTGGGCATCAGCTACGGCCGCAAGAAGCGCCGCCAGCGCCGCCGCGCCCCCCCC  
GACAGCGAGGTGCACCAGGTGAGCCTGCCCAAGCAGCCCGCCAGCCAGCCCCAGGGCGA  
CCCCACCGGCCCAAGGAGAGCAAGAAGAAGGTGGAGCGCGAGACCGAGACCGACCCCG  
TGCCTAG

**FIG. 79**

(SEQ ID NO:88)

*SUBSTITUTE SHEET (RULE 26)*





127 / 131

## Alignment GagMod vs GP1\_GP2

										Section 6
										456
GagMod.SF2	(381)	381	390	400	410	420	430	440		
GagProtMod.SF2 (GP1)	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACTGAGGGCCAGATGGTGCAACAGGCCATCAGCCCCCGC								
GagProtMod.SF2 (GP2)	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACTGAGGGCCAGATGGTGCAACAGGCCATCAGCCCCCGC								
Consensus	(381)	CAGCCAGGTGAGCCAGAACTACCCCATCGTGAGAACTGAGGGCCAGATGGTGCAACAGGCCATCAGCCCCCGC								
										Section 7
										532
GagMod.SF2	(457)	457	470	480	490	500	510	520		
GagProtMod.SF2 (GP1)	(457)	ACCCTGAACGCTGGGTGAGGTGGTGGAGGAGAGGCTTCAGCCCCAGGTGATCCCCATGTTTCAGCGCCCTGA								
GagProtMod.SF2 (GP2)	(457)	ACCCTGAACGCTGGGTGAGGTGGTGGAGGAGAGGCTTCAGCCCCAGGTGATCCCCATGTTTCAGCGCCCTGA								
Consensus	(457)	ACCCTGAACGCTGGGTGAGGTGGTGGAGGAGAGGCTTCAGCCCCAGGTGATCCCCATGTTTCAGCGCCCTGA								
										Section 8
										608
GagMod.SF2	(533)	533	540	550	560	570	580	590		
GagProtMod.SF2 (GP1)	(533)	GCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCCGTGGCGGGCCACCAGGCCGCCATGCAGATGCT								
GagProtMod.SF2 (GP2)	(533)	GCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCCGTGGCGGGCCACCAGGCCGCCATGCAGATGCT								
Consensus	(533)	GCGAGGGCGCCACCCCCAGGACCTGAACACGATGTTGAACACCCGTGGCGGGCCACCAGGCCGCCATGCAGATGCT								
										Section 9
										684
GagMod.SF2	(609)	609	620	630	640	650	660	670		
GagProtMod.SF2 (GP1)	(609)	GAAGGAGACCATCAACGAGGAGGCGCCCGCGAGTGGGACCCGCTGCACCCCGTGCACGGCGGCCCATCGCCCCCGGC								
GagProtMod.SF2 (GP2)	(609)	GAAGGAGACCATCAACGAGGAGGCGCCCGCGAGTGGGACCCGCTGCACCCCGTGCACGGCGGCCCATCGCCCCCGGC								
Consensus	(609)	GAAGGAGACCATCAACGAGGAGGCGCCCGCGAGTGGGACCCGCTGCACCCCGTGCACGGCGGCCCATCGCCCCCGGC								
										Section 10
										760
GagMod.SF2	(685)	685	690	700	710	720	730	740	750	
GagProtMod.SF2 (GP1)	(685)	CAGATGCGGAGAGCCCGCGGCGAGCGACATCGCCGGCACCCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCA								
GagProtMod.SF2 (GP2)	(685)	CAGATGCGGAGAGCCCGCGGCGAGCGACATCGCCGGCACCCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCA								
Consensus	(685)	CAGATGCGGAGAGCCCGCGGCGAGCGACATCGCCGGCACCCAGCACCCCTGCAGGAGCAGATCGGCTGGATGACCA								

FIG. 80B

SUBSTITUTE SHEET (RULE 26)



129 / 131

# Alignment GagMod vs GP1\_GP2

[illegible]

**FIG. 80D**



131/131

TataminoSF162.opt

ATGGAGCCCCGTGGACCCCCCGCCCTGGAGCCCTGGAGCGCAACCCCGGCAGCCAGCCCCAA  
GACCGCCCTGGACCAACTGCTACTGCAAGAGAGTGCTGCTTCCACTGCCAGGTGTGCTT  
CATCACCAAGGGCCTGGGCATCAGCTACGGCCGCAAGAGCGCCGCCAGCGCCGC

**FIG. 81**  
(SEQ ID NO:89)

Tat\_Cys22\_SF162

MEPVDPRLEPWKHPSQPKTAGTNCYCKKCCFHCQVCFITKGLGISYGRKKRRRRRAPPDSE  
VHQVSLPKQPASQPQGDPTGPKESKKKVERETETDPVHZ

**FIG. 82**  
(SEQ ID NO:90)

SUBSTITUTE SHEET (RULE 26)

## SEQUENCE LISTING

&lt;110&gt; Chiron Corporation

<120> IMPROVED EXPRESSION OF HIV POLYPEPTIDES AND PRODUCTION  
OF VIRUS-LIKE PARTICLES

&lt;130&gt; 1621.100

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 90

&lt;170&gt; PatentIn Ver. 2.0

&lt;210&gt; 1

&lt;211&gt; 1509

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 1

```

atgggtgcga gagcgctcggg attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcagggag 120
ctagaacgat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtagcaa cctctattg tgtacatcaa aggatagatg taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaaacag cagccaggct agccaaaatt accctatagt gcagaaccta 420
caggggcaaa tgggtacatca ggccatatca cctagaactt taaatgcatg ggtaaaagta 480
gtagaagaaa aggcctttcag ccagaagta ataccatgt tttcagcatt atcagaagga 540
gccacccac aagattttaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaattgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
catgcagggc ctattgcacc aggcctaatg agagaaccaa ggggaagtga catagcagga 720
actactagta ccttcagga acaaatagga tggatgacaa ataatccacc tatcccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaac cctttagaga ttatgtagac 900
cgtttctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgt tgggtccaaa tgcaaaccca gattgttaaga ctatttttaa agcattggga 1020
ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg acccgccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaagaaag actgttaagt gtttcaattg tggcaaagaa 1200
gggcacatag ccaaaaattg cagggcccct aggaaaaagg gctgttgagg atgtggaagg 1260
gaaggacacc aaatgaaaga ttgcaactgag agacaggcta attttttagg gaagatctgg 1320
ccttcctaca aggggaaggcc aggggaatttt cttcagagca gaccagagcc aacagcccca 1380
ccagaagaga gcttcagggt tggggaggag aaaacaactc cctctcagaa gcaggagccg 1440
atagacaagg aactgtatcc ttttaacttc ctcagatcac tctttggcaa cgaccctcgg 1500
tcacaataa
1509

```

&lt;210&gt; 2

&lt;211&gt; 1845

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 2

```

atgggtgcga gagcgctcggg attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcagggag 120
ctagaacgat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180

```

```

ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtagcaa ccctctattg tgtacatcaa aggatagatg taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaaacag cagccaggtc agccaaaatt accctatagt gcagaaccta 420
caggggcaaaa tggtagcatc ggccatatca cctagaactt taaatgcatg ggtaaaagta 480
gtagaagaaa aggcctttcag cccagaagta atacccatgt tttcagcatt atcagaagga 540
gccacccccc aagattttaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
catgcagggc ctattgcacc aggccaaatg agagaaccaa ggggaagtga catagcagga 720
actactagta cccttcagga acaaataagga tggatgacaa ataatccacc tatcccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaac cctttagaga ttatgtagac 900
cggttctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgt tgggtccaaaa tgcaaaccca gattgtaaga ctatttttaa agcattggga 1020
ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg acccggccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaagaaag actgttaagt gtttcaattg tggcaaaaga 1200
gggcacatag ccaaaaattg cagggccctt aggaaaaagg gctgttggag atgtggaagg 1260
gaaggacacc aaatgaaaga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
ccttctaca agggaaggcc agggaatttt cttcagagca gaccagagcc aacagcccca 1380
ccagaagaga gcttcagggt tggggaggag aaaacaactc cctctcagaa gcaggagccg 1440
atagacaagg aactgtatcc ttttaacttc ctcagatcac tctttggcaa cgaccctcgt 1500
tcacaataag gatagggggg caactaaagg agctctatt agatacagga gcagatgata 1560
cagtattaga agaaatgaat ttgccaggaa aatggaaacc aaaaatgata gggggaattg 1620
gaggttttat caaagtaaga cagtacgatc agatacctgt agaaatctgt ggacataaag 1680
ctataggtac agtattagta ggacctacac ctgtcaacat aattggaaga aatctgttga 1740
ctcagattgg ttgtacttta aatttcccca ttagtcttat tgaaactgta ccagtaaaat 1800
taaaagccag aatggatggc ccaaagttta agcaatggcc attga 1845

```

&lt;210&gt; 3

&lt;211&gt; 4313

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 3

```

atgggtgcga gagegtcggt attagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcaggag 120
ctagaacgat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtagcaa ccctctattg tgtacatcaa aggatagatg taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaaacag cagccaggtc agccaaaatt accctatagt gcagaaccta 420
caggggcaaaa tggtagcatc ggccatatca cctagaactt taaatgcatg ggtaaaagta 480
gtagaagaaa aggcctttcag cccagaagta atacccatgt tttcagcatt atcagaagga 540
gccacccccc aagattttaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaatgttaa aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagtg 660
catgcagggc ctattgcacc aggccaaatg agagaaccaa ggggaagtga catagcagga 720
actactagta cccttcagga acaaataagga tggatgacaa ataatccacc tatcccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaac cctttagaga ttatgtagac 900
cggttctata aaactctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgt tgggtccaaaa tgcaaaccca gattgtaaga ctatttttaa agcattggga 1020
ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg acccggccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaagaaag actgttaagt gtttcaattg tggcaaaaga 1200
gggcacatag ccaaaaattg cagggccctt aggaaaaagg gctgttggag atgtggaagg 1260
gaaggacacc aaatgaaaga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
ccttctaca agggaaggcc agggaatttt cttcagagca gaccagagcc aacagcccca 1380
ccagaagaga gcttcagggt tggggaggag aaaacaactc cctctcagaa gcaggagccg 1440

```

```

atagacaagg aactgtatcc tttaacttcc ctcatatcac tctttggcaa cgaccctctg 1500
tcacaataag gatagggggg caactaaagg aagctctatt agatacagga gcagatgata 1560
cagtattaga agaaatgaat ttgccaggaa aatggaaacc aaaaatgata gggggaattg 1620
gagggttttat caaagtgaag cagtacgac agatacctgt agaaatctgt ggacataaag 1680
ctataggtag agtattagta ggacctacac ctgtcaacat aattggaaga aatctgttga 1740
ctcagattgg ttgtacttta aatttcccca ttagtccat tgaaactgta ccagtaaaat 1800
taaagccagg aatggatggc ccaaaagtta agcaatggcc attgacagaa gaaaaataa 1860
aagcattagt agagatatgt acagaaatgg aaaaggaagg gaaaatttca aaaattgggc 1920
ctgaaaatcc atacaatact ccagtatttg ctataaagaa aaaagacagt actaaatgga 1980
gaaaactagt agatttcaga gaacttaata aaagaactca agacttctgg gaagttcagt 2040
taggaatacc acaccccgca ggggttaaaa agaaaaaatc agtaacagta ttggatgtgg 2100
gtgatgcata cttttcagtt cccttagata aagactttag aaagtatact gcatttacca 2160
tacctagtag aaacaatgag acaccaggga ttagatatca gtacaatgtg ctgccacagg 2220
gatggaaagg atcaccagca atattccaaa gtagcatgac aaaaatctta gagcctttta 2280
gaaaacagaa tccagacata gttatctatc aatacatgga tgatttgtat gtaggatctg 2340
acttagaaat agggcagcat agaacaaaaa tagaggaact gagacagcat ctgttgaggt 2400
ggggatttac cacaccagac aaaaaacatc agaaaagaacc tccattcctt tggatgggtt 2460
atgaactcca tcctgataaa tggcagtag agcctataat gctgccagaa aaagacagct 2520
ggactgtcaa tgacatacag aagttagtgg gaaaattgaa ttgggcaagt cagatttatg 2580
cagggattaa agtaaaagcag ttatgtaaac tccttagagg aaccaaagca ctaacagaag 2640
taataccact aacagaagaa gcagagctag aactggcaga aaacaggagg attctaaaag 2700
aaccagtaca tgaagtatac tatgacccat caaaagactt agtagcagaa atacagaagc 2760
aggggcaagg ccaatggaca tatcaaatat tccttagagg attaaaaaat ctgaaaacag 2820
gaaagtatgc aaggatgagg ggtgccaca ctaatgatgt aaaacagtta acagaggcag 2880
tgcaaaaagt atccacagaa agcatagtaa tatggggaaa gattcctaaa tttaaactac 2940
ccatacaaaa ggaacatgg gaagcatggt ggtaggagta ttggcaagct acctggattc 3000
ctgagtggga gtttgtcaat acccctccct tagtgaaatt atggtaccag ttagagaaag 3060
aaccctatag aggagcagaa actttctatg tagatggggc agctaatagg gagactaaat 3120
taggaaaagc aggatatggt actgacagag gaagacaaaa agttgtctcc atagctgaca 3180
caacaaatca gaagactgaa ttacaagcaa ttcactagc tttgcaggat tcgggattag 3240
aagtaaacat agtaacagac tcacaatatg cattaggaat cattcaagca caaccagata 3300
agagtgaatc agagttagtc agtcaaataa tagagcagtt aataaaaaag gaaaagggtc 3360
acctggcatg ggtaccagca ccaaaaggaa ttggaggaaa tgaacaagta gataaattag 3420
tcagtgtcgg aatcaggaaa gtactatttt tgaatggaat agataaggcc- caagaagaac 3480
atgagaaata tcacagtaat tggagagcaa tggctagtga ttttaacctg ccacctgtag 3540
tagcaaaaaga aatagtagcc agctgtgata aatgtcagct aaaaggagaa gccatgcatg 3600
gacaagttaga ctgtagtcca ggaatatggc aactagattg tacacatcta gaaggaaaa 3660
ttatcctggt agcagttcat gtagccagtg gatatataga agcagaagtt attccagcag 3720
agacagggca ggaacagca tattttctct taaaattagc aggaagatgg ccagtaaaaa 3780
caatacatag agacaatggc agcaatttca ccagtactac ggttaaggcc gcctgttgg 3840
gggcagggat caagcaggaa tttggcattc cctacaatcc ccaaagtcaa ggagtagtag 3900
aatctatgaa taatgaatta aagaaaatta taggacaggt aagagatcag gctgaacacc 3960
ttaagacagc agtacaatg gcagtatcca tccacaattt taaaagaaaa ggggggattg 4020
ggggatagag tgcaggggaa agaattagtag acataatagc aacagacata caaactaaag 4080
aactacaaaa gcaattaca aaaattcaaa attttcgggt ttattacagg gacaacaaag 4140
atcccccttg gaaaggacca gcaaagcttc tctggaaagg tgaaggggca gtagtaatac 4200
aagataatag tgacataaaa gtatgccaa gaagaaaagc aaaaatcatt agggattatg 4260
gaaaacagat ggcagggtgag gattgtgtgg caagtagaca ggatgaggat tag 4313

```

&lt;210&gt; 4

&lt;211&gt; 1515

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic  
HIV-Gag

&lt;400&gt; 4



```

gccacccatgg gcgcccgcgc cagcgtgctg agcggcgggcg agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggcccagc 120
cgcgagctgg agcgtctcgc cgtgaacccc ggccctgctgg agaccagcga gggctgccgc 180
cagatccctgg gccagctgca gcccagcctg cagaccggca gcgaggagct gcgcagcctg 240
tacaacaccg tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcgcgcg ccggcaccgg caacagcagc caggtgagcc agaactaccc catcgtgcag 420
aacctgcagg gccagatggg gcaccaggcc atcagcccc gcaccctgaa cgccctgggtg 480
aagggtgggtg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cgccctgagc 540
gaggggcgcca cccccagga cctgaacacg atgttgaaca ccgtggggcg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgctgcac 660
cccgtgcacg ccggccccat cgcccccggc cagatgcgcg agccccgcgg cagcgacatc 720
gccggcacca ccagcaccct gcaggagcag atcggctgga tgaccaacaa cccccccatc 780
cccgtggggcg agatctacaa gcggtggatc atcctggggc tgaacaagat cgtgcggatg 840
tacagcccca ccagcatcct ggacatccgc cagggcccca aggagccctt ccgcgactac 900
gtggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
atgaccgaga ccctgctggt gcagaacgcc aaccccgact gcaagaccat cctgaaggct 1020
ctcggccccg cggccaccct ggaggagatg atgaccgcct gccaggggcg gggcgggccc 1080
ggccacaagg ccgcgtgct ggccgaggcg atgagccagg tgacgaacc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cggaagaccg tcaagtgtt caactgcggc 1200
aaggaggggc acaccgcag gaactgccgc gcccccgca agaagggtg ctggcgctgc 1260
ggccgcgagg gccaccagat gaaggactgc accgagcgcc agggcaactt cctgggcaag 1320
atctggccca gctacaagg ccgccccggc aacttcctgc agagccgccc cgagcccacc 1380
gcccccccg aggagagctt ccgcttcggc gaggagaaga ccacccccag ccagaagcag 1440
gagcccatcg acaaggagct gtacccctg accagcctgc gcagcctgtt cggcaacgac 1500
cccagcagcc agtaa

```

1515

&lt;210&gt; 5

&lt;211&gt; 1853

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic  
HIV-Gag-protease

&lt;400&gt; 5

```

gccacccatgg gcgcccgcgc cagcgtgctg agcggcgggcg agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggcccagc 120
cgcgagctgg agcgtctcgc cgtgaacccc ggccctgctgg agaccagcga gggctgccgc 180
cagatccctgg gccagctgca gcccagcctg cagaccggca gcgaggagct gcgcagcctg 240
tacaacaccg tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcgcgcg ccggcaccgg caacagcagc caggtgagcc agaactaccc catcgtgcag 420
aacctgcagg gccagatggg gcaccaggcc atcagcccc gcaccctgaa cgccctgggtg 480
aagggtgggtg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cgccctgagc 540
gaggggcgcca cccccagga cctgaacacg atgttgaaca ccgtggggcg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgctgcac 660
cccgtgcacg ccggccccat cgcccccggc cagatgcgcg agccccgcgg cagcgacatc 720
gccggcacca ccagcaccct gcaggagcag atcggctgga tgaccaacaa cccccccatc 780
cccgtggggcg agatctacaa gcggtggatc atcctggggc tgaacaagat cgtgcggatg 840
tacagcccca ccagcatcct ggacatccgc cagggcccca aggagccctt ccgcgactac 900
gtggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
atgaccgaga ccctgctggt gcagaacgcc aaccccgact gcaagaccat cctgaaggct 1020
ctcggccccg cggccaccct ggaggagatg atgaccgcct gccaggggcg gggcgggccc 1080
ggccacaagg ccgcgtgct ggccgaggcg atgagccagg tgacgaacc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cggaagaccg tcaagtgtt caactgcggc 1200
aaggaggggc acaccgcag gaactgccgc gcccccgca agaagggtg ctggcgctgc 1260
ggccgcgaag gacaccaaag gaaagattgc actgagagac aggctaattt ttagggaag 1320

```

```

atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
gccccaccag aagagagctt cagggtttggg gaggagaaaa caactccctc tcagaagcag 1440
gagccgatat acaaggaact gtatccttta acttccctca gatcactctt tggcaacgac 1500
ccctcgtcac agtaaggatc ggcgccagc tcaaggaggc gctgctcgac accggcgccg 1560
acgacaccgt gctggaggag atgaacctgc cgggcaagtg gaagcccaag atgatcggcg 1620
ggatcggggg cttcatcaag gtgcggcagt acgaccagat ccccgaggag atctgcggcc 1680
acaaggccat cggcaccgtg ctggtggggc ccacccccgt gaacatcatc ggccgcaacc 1740
tgctgaccca gatcggctgc accctgaact tccccatcag ccccatcgag acggtgcccc 1800
tgaagctgaa gccggggatg gacggcccca aggtcaagca gtggcccctg taa 1853

```

&lt;210&gt; 6

&lt;211&gt; 4319

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: synthetic

HIV-Gag-polymerase

&lt;400&gt; 6

```

gccaccatgg ggcggcgcg cagcgtgctg agcggcgggc agctggacaa gtggggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggcccagc 120
cgcgagctgg agcgtctcgc cgtgaacccc ggctgctgg agaccagcga gggctgcccgc 180
cagatcctgg gccagctgca gccagcctg cagaccggca gcgaggagct gcgcagcctg 240
tacaacaccg tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggccctgg agaagctga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcggcgg cggcaccgg caacagcagc caggtgagcc agaactacc catcgtgcag 420
aacctgcagg gccagatggt gcaccaggcc atcagcccc gcaccctgaa cgcctgggtg 480
aaggtggtgg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cgccctgagc 540
gagggcgcca cccccagga cctgaacacg atgttgaaca ccgtggcgcg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggcgg cagagtggga ccgctgac 660
cccgctcacg cggcccccac cgccccggc cagatcgcg agccccggc cagcgacatc 720
gccggcacca ccagcaccct gcaggagcag atcggctgga tgaccaacaa ccccccatc 780
cccgctggcg agatctacaa gcggtggatc atcctgggcc tgaacaagat cgtcgggatg 840
tacagcccca ccagcatcct ggacatccgc cagggcccca aggagccctt ccgcgactac 900
gtggaccgct tctacaagac cctgcgcgt cgcaggcca gccaggacgt gaagaactgg 960
atgaccgaga cctgctggt gcagaacgcc aaccgcgact gcaagaccat cctgaaggct 1020
ctcgcccccg cggccaccct ggaggagatg atgaccgct gccaggcggt gggcgccccc 1080
ggccacaagg cccgcgtgct ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
atgatgcagc gcggaactt ccgcaaccag cggaaagacc tcaagtgtt caactgcggc 1200
aaggagggcc acaccgccg gaactgccgc gcccccgca agaaggcgct ctggcgctgc 1260
ggccgcgaag gacaccaa atgaaattgc actgagagac aggtctaatt tttagggaag 1320
atctggcctt cctacaaggg aaggccaggg aattttcttc agagcagacc agagccaaca 1380
gccccaccag aagagagctt cagggtttggg gaggagaaaa caactccctc tcagaagcag 1440
gagccgatat acaaggaact gtatccttta acttccctca gatcactctt tggcaacgac 1500
ccctcgtcac agtaaggatc ggcgccagc tcaaggaggc gctgctcgac accggcgccg 1560
acgacaccgt gctggaggag atgaacctgc cgggcaagtg gaagcccaag atgatcggcg 1620
ggatcggggg cttcatcaag gtgcggcagt acgaccagat ccccgaggag atctcggcg 1680
acaaggccat cggcaccgtg ctggtggggc ccacccccgt gaacatcatc ggccgcaacc 1740
tgctgaccca gatcggctgc accctgaact tccccatcag ccccatcgag acggtgcccc 1800
tgaagctgaa gccggggatg gacggcccca aggtcaagca gtggcccctg accgaggaga 1860
agatcaaggc cctggtggag atctgcaccg agatggagaa ggagggcaag atcagcaaga 1920
tcggccccga gaacccctac aacacccccg tgttcgccat caagaagaag gacagacca 1980
agtggcgcaa gctggtggac ttccgcgagc tgaacaagcg caccaggac ttctgggagg 2040
tgcatctggg catccccac cccgcggcc tgaagaagaa gaagagcgtg accgtgctgg 2100
acgtgggcga gcctacttc agcgtgcccc tggacaagga cttccgcaag tacaccgct 2160
tcaccatccc cagcatcaac aacgagaccc cggcatccg ctaccagtac aacgtgctgc 2220
cccagggtg gaaggcgag cccgccatct tccagagcag catgaccaag atcctggagc 2280
ccttccgcaa gcagaacccc gacatcgtga tctaccagta catggacgac ctgtacgtg 2340

```

```

gcagcgacct ggagatcgcc cagcaccgca ccaagatcga ggagctgcgc cagcacctgc 2400
tgcgctgggg cttcaccacc cccgacaaga agcaccagaa ggagccccc ttcctgtgga 2460
tgggctacga gctgcacccc gacaagtggg cctgtcagcc catcatgctg cccgagaagg 2520
acagctggac cgtgaacgac atccagaagc tgggtgggcaa gctgaactgg gccagccaga 2580
tctacgccgg catcaagggtg aagcagctgt gcaagctgct gcgcggcacc aaggccctga 2640
ccgaggtgat cccctgacc gaggaggccg agctggagct ggccgagaac cgcgagatcc 2700
tgaaggagcc cgtgcacgag gtgtactacg accccagcaa ggacctggtg gccgagatcc 2760
agaagcaggg ccagggccag tggacctacc agatctacca ggagcccttc aagaacctga 2820
agaccggcaa gtacgcccg c atgcgcgccg cccacaccaa cgacgtgaag cagctgaccg 2880
aggccgtgca gaaggtgagc accgagagca tctgtatctg gggcaagatc cccaagtcca 2940
agctgcccac ccagaaggag acctgggagg cctgggtgat ggagtactgg caggccacct 3000
ggatccccga gtgggagttc gtgaacaccc ccccccgtgt gaagctgtgg taccagctgg 3060
agaaggagcc catcgtgggc gccgagacct tctacgtgga cggcgccgcc aaccgcgaga 3120
ccaagctggg caaggccggc tacgtgaccg acccgcgccg ccagaagggtg gtgagcatcg 3180
ccgacaccac caaccagaag accgagctgc agggcatcca cctggccctg caggacagcg 3240
gcctggaggt gaacatcgtg accgacagcc agtacgcctt gggcatcatc caggcccagc 3300
ccgacaagag cgagagcgag ctggtgagcc agatcatcga gcagctgatc aagaaggaga 3360
aggtgtacct ggccctgggtg cccgcccaca agggcatcgg cggcaacgag caggtggaca 3420
agctggtgag cgccggcatc cgcaagggtg tgttcctgaa cggcatcgac aaggcccagg 3480
aggagcacga gaagtaccac agcaactggc gcgccatggc cagcgacttc aacctgcccc 3540
ccgtggtggc caaggagatc gtggccagct gcgacaagtg ccagctgaag ggcgaggcca 3600
tgcacggcca ggtggactgc agccccggca tctggcagct ggactgcacc cacctggagg 3660
gcaagatcat cctggtggcc gtgcacgtgg ccagcggtca catcgaggcc gaggtgatcc 3720
ccgcccagac cggccaggag accgcctact tctgtctgaa gctggccggc cgtggtggccg 3780
tgaagaccat ccacaccgac aacggcagca acttcaccag caccaccgtg aaggccgctt 3840
gctggtgggc cggcatcaag caggagtctg gcacccccta caaccccag agccaggggc 3900
tgggtggagag catgaacaac gagctgaaga agatcatcgg ccaggtgcgc gaccaggccg 3960
agcacctgaa gaccgcccgtg cagatggccg tgttcaccca caacttcaag cgcaaggggc 4020
gcatcgccgg ctacagcgcc ggcgagcgca tctgtggacat catcgccacc gacatccaga 4080
ccaaggagct gcagaagcag atcaccaaga tccagaactt ccgctgttac taccgcgaca 4140
acaaggaccc cctgtggaag ggccccggca agctgctgtg gaaggcgag ggcgcccgtg 4200
tgatccagga caacagcgac atcaagggtg tgccccggcg caaggccaag atcatccgcg 4260
actacggcaa gcagatggcc ggcgacgact gcgtggccag ccgcccaggac gaggactag 4319

```

&lt;210&gt; 7

&lt;211&gt; 2031

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic  
HIV-Gag/HCV-core fusion polypeptide

&lt;400&gt; 7

```

gccaccatgg gcgcccgcgc cagcgtgctg agcggcgccg agctggacaa gtgggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtggggccagc 120
cgcgagctgg agcgtcttcg cgtgaacccc ggccctgctg agaccagcga gggctgccgc 180
cagatcctgg gccagctgca gccagcctg cagaccggca gcgaggagct gcgcagcctg 240
tacaacaccg tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcccggc ccggcaccgg caacagcagc caggtgagcc agaactacc catcgtgcag 420
aacctgcagg gccagatggt gcaccaggcc atcagcccc gcacctgaa cgcctgggtg 480
aagggtggtg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cgccctgagc 540
gagggcgcca cccccaggga cctgaacacg atgttgaaca ccgtgggcgg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgctgcac 660
ccgtgacag ccggcccat gcgccccggc cagatgcgcg agccccggcg cagcgacatc 720
gccggcacca ccagaccct gcaggagcag atcggtgga tgaccaacaa ccccccatc 780
cccggtggcg agatctacaa gcggtggatc atcctgggcc tgaacaagat cgtgcggatg 840
tacagcccca ccagcatcct ggacatccgc cagggcccca aggagccctt ccgcgactac 900

```

```

gtggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 96)
atgaccgaga ccctgctggt gcagaacgcc aaccccgact gcaagaccat cctgaaggct 1020
ctcgcccccg cgccaccctt ggaggagatg atgaccgcct gccaggcgct gggcgccccc 1080
ggccacaagg ccgcgtgctt ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cggaagaccg tcaagtgett caactgcggc 1200
aaggaggggc acaccgccag gaactgccgc gccccccgca agaagggtgt ctggcgctgc 1260
ggccgcgagg gccaccagat gaaggactgc accgagcgcc agggcaactt cctgggcaag 1320
atctggccca gctacaaggg ccgcccgggc aacttcctgc agagccgccc cgagcccacc 1380
gcccccccg aggagagctt ccgcttcggc gaggagaaga ccacccccag ccagaagcag 1440
gagcccatcg acaaggagct gtacccccctg accagcctgc gcagcctgtt cggcaacgac 1500
cccagcagcc agtcgacgaa tcctaaacct caaagaaaaa acaaacgtaa caccaacctg 1560
cgccacagg acgtcaagtt ccgggtgggc ggtcagatcg ttggtggagt ttacttgttg 1620
ccgcgcaggg gccctagatt gggtgtgcgc gcgacgagaa agacttcga gcggtcgcaa 1680
cctcgaggta gacgtcagcc tatcccaag gctcgtcggc ccgagggcag gacctgggt 1740
cagcccggtt acccttggcc cctctatggc aatgagggtt gcgggtgggc gggatggctc 1800
ctgtctcccc gtggctctcg gcctagctgg ggcgccacag acccccgcg taggtcgcg 1860
aatttgggtt aggtcatcga tacctttacg tgcggcttcg ccgacctcat ggggtacata 1920
ccgctcgtcg gcgcccctct tggaggcgct gccaggggcc tggcgcatgg cgtccgggtt 1980
ctggaagacg gcgtgaacta tgcaacaggg aaccttcctg gttgctctta g 2031

```

&lt;210&gt; 8

&lt;211&gt; 2025

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic  
HIV-Gag/HCV-Core fusion polypeptide

&lt;400&gt; 8

```

atgggtgcga gagcgtcggt attaagcggg ggagaattag ataaatggga aaaaattcgg 60
ttaaggccag ggggaaagaa aaaatataag ttaaaacata tagtatgggc aagcagggag 120
ctagaacgat tcgcagtcaa tcctggcctg ttagaaacat cagaaggctg cagacaaata 180
ttgggacagc tacagccatc ccttcagaca ggatcagaag aacttagatc attatataat 240
acagtagcaa ccctctattg tgtacatcaa aggatagatg taaaagacac caaggaagct 300
ttagagaaga tagaggaaga gcaaaacaaa agtaagaaaa aggcacagca agcagcagct 360
gcagctggca caggaaacag cagccaggtc agccaaaatt accctatagt gcagaacctt 420
cagggggcaaa tggtagatca ggccatatca cctagaactt taaatgcatt ggtaaaagta 480
gtagaagaaa aggetttcag ccagaaagta ataccatgtt ttccagcatt atcagaagga 540
gccacccac aagatttaaa caccatgcta aacacagtgg ggggacatca agcagccatg 600
caaagtgtta aagagactat caatgaggaa gctgcagaat gggatagagt gcatccagt 660
catgcagggc ctattgcacc aggccaaatg agagaaccaa ggggaagtga catagcagga 720
actactagta cccttcagga acaaatagga tggatgacaa ataatccacc tatccagta 780
ggagaaatct ataaaagatg gataatcctg ggattaaata aaatagtaag aatgtatagc 840
cctaccagca ttctggacat aagacaagga ccaaaggaac ccttagaga ttatgtagac 900
cggttctata aaactcctaag agccgaacaa gcttcacagg atgtaaaaaa ttggatgaca 960
gaaaccttgt tggccaaaaa tgcaaaccca gattgtaaga ctatttttaa agcattggga 1020
ccagcagcta cactagaaga aatgatgaca gcatgtcagg gagtgggggg acccgccat 1080
aaagcaagag ttttggctga agccatgagc caagtaacaa atccagctaa cataatgatg 1140
cagagaggca attttaggaa ccaaagaaag actgttaagt gtttcaattg tggcaagaa 1200
gggcacatag ccaaaaattg cagggccctt aggaaaaagg gctgttggag atgtggaagg 1260
gaaggacacc aaatgaaaga ttgcactgag agacaggcta attttttagg gaagatctgg 1320
ccttcctaca agggaaggcc agggaatttt cttcagagca gaccagagcc aacagcccca 1380
ccagaagaga gcttcagggt tggggaggag aaaacaactc ccttcagaa gcaggagccg 1440
atagacaagg aactgtatcc ttttaacttc ctcagatcac tctttggcaa cgacccctcg 1500
tcacagtcga cgaatcctaa acctcaaaga aaaaacaaac gtaacaccaa ccgtcgccca 1560
caggacgtca agttcccggtg tggcggtcag atcgttgggt gagtttactt gttgccgcgc 1620
agggggcccta gattgggtgt gcgcgcgacg agaaagactt ccgagcggtc gcaacctcga 1680
ggtagacgtc agcctatccc caaggctcgt cggcccagg gcaggacctg ggctcagccc 1740

```

```

gggtaccctt ggccctctta tggcaatgag ggctgcgggt gggcgggatg gctcctgtct 1800
ccccgtggt ctcggcctag ctggggcccc acagaccccc ggcgtaggtc gcgcaatttg 1860
ggtaaggta tcgataccct tacgtgcggc ttccggcacc tcatgggta cataccgtc 1920
gtcggcgccc ctcttggagg cgtgccagg gccctggcgc atggcgccg ggttctggaa 1980
gacggcgtga actatgcaac agggaaacct cctggttgct cttag 2025

```

&lt;210&gt; 9

&lt;211&gt; 1268

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic Gag  
common region

&lt;400&gt; 9

```

gccaccatgg ggcggcgcg cagcgtgctg agcggcgggc agctggacaa gtggggagaag 60
atccgcctgc gccccggcgg caagaagaag tacaagctga agcacatcgt gtgggcccagc 120
cgcgagctgg agcgcttcgc cgtgaacccc ggctgctgg agaccagcga gggctgccgc 180
cagatcctgg gccagctgca gcccagcctg cagaccggca gcgaggagct gcgcagcctg 240
tacaacaccg tggccaccct gtactgcgtg caccagcgca tcgacgtcaa ggacaccaag 300
gaggccctgg agaagatcga ggaggagcag aacaagtcca agaagaaggc ccagcaggcc 360
gccgcccggc ccggcaccgg caacagcagc caggtgagcc agaactacc catcgtgcag 420
aacctgcagg gccagatggg gcaccaggcc atcagcccc gcacctgaa cgcttgggtg 480
aaggtggtgg aggagaaggc cttcagcccc gaggtgatcc ccatgttcag cgccctgagc 540
gagggcgcca cccccagga cctgaacacg atgttgaaca ccgtggggcg ccaccaggcc 600
gccatgcaga tgctgaagga gaccatcaac gaggaggccg ccgagtggga ccgctgcac 660
cccgtgcacg ccggccccc catcggccggc cagatgcgcg agccccggc cagcgacatc 720
gccggcacca ccagcaccct gcaggagcag atcggctgga tgaccaacaa ccccccatc 780
cccgtggcg agatctacaa gcggtggatc atcctgggccc tgaacaagat cgtgcggatg 840
tacagcccca ccagcatcct ggacatccgc cagggcccca aggagccct ccgcgactac 900
gtggaccgct tctacaagac cctgcgcgct gagcaggcca gccaggacgt gaagaactgg 960
atgaccgaga ccctgctggg gcagaacgcc aaccccgaact gcaagaccat cctgaaggct 1020
ctcgccccc cgccaccct ggaggagatg atgaccgct gccaggcggt gggcgggccc 1080
ggccacaagg ccgcgtgct ggccgaggcg atgagccagg tgacgaaccc ggcgaccatc 1140
atgatgcagc gcggcaactt ccgcaaccag cggaagaccg tcaagtgctt caactgcggc 1200
aaggagggcc acaccgccag gaactgccgc gcccccgca agaagggctg ctggcgctgc 1260
ggccgcga 1268

```

&lt;210&gt; 10

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: HIV-Gag  
peptide p7G

&lt;400&gt; 10

```

Gly Gly His Gln Ala Ala Met Gln Met Leu Lys Glu Thr Ile Asn Glu
  1             5             10             15

```

```

Glu Ala Ala Glu
      20

```

&lt;210&gt; 11

&lt;211&gt; 30

&lt;212&gt; DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer GAG5

<400> 11

aagaattcca tgggtgcgag agcgtcggtta

30

<210> 12

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer  
p55-SAL3

<400> 12

attcgtcgac tgtgacgagg ggtcgttgcc

30

<210> 13

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer  
CORESAL5

<400> 13

atttgcgac gaatcctaaa cctcaaagaa aaac

34

<210> 14

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer 173CORE

<400> 14

tattggatcc taagagcaac caggaagggtt c

31

<210> 15

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer MS65

<400> 15

cgaccatcat ggatgcagcg c

21

<210> 16

<211> 30

<212> DNA

<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: primer MS66

&lt;400&gt; 16

aggattcgtc gagtcgctgc tggggctcgtt

30

&lt;210&gt; 17

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: primer XPANXNF

&lt;400&gt; 17

gcacgtgggc ccggcgcttc tagagc

26

&lt;210&gt; 18

&lt;211&gt; 26

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: primer XPANXNR

&lt;400&gt; 18

gctctagagg cgccgggccc acgtgc

26

&lt;210&gt; 19

&lt;211&gt; 20

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: HIV p55 Gag  
Major Homology Region

&lt;400&gt; 19

Asp Ile Arg Gln Gly Pro Lys Glu Pro Phe Arg Asp Tyr Val Asp Arg  
1 5 10 15

Phe Tyr Lys Thr

20

&lt;210&gt; 20

&lt;211&gt; 60

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic p55  
Gag Major Homology Region

&lt;400&gt; 20

gacatccgcc agggccccaaggagcccttc cgcgactacg tggaccgctt ctacaagacc 60

&lt;210&gt; 21

&lt;211&gt; 15

<212> PRT  
<213> Human immunodeficiency virus

<400> 21  
Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg  
1 5 10 15

<210> 22  
<211> 5  
<212> PRT  
<213> Human immunodeficiency virus

<400> 22  
Lys Ala Lys Arg Arg  
1 5

<210> 23  
<211> 4  
<212> PRT  
<213> Human immunodeficiency virus

<400> 23  
Arg Glu Lys Arg  
1

<210> 24  
<211> 15  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: aa of  
mut7.SF162 cleavage site

<400> 24  
Ala Pro Thr Lys Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser  
1 5 10 15

<210> 25  
<211> 15  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: aa of  
mut8.SF162 cleavage site

<400> 25  
Ala Pro Thr Ile Ala Ile Ser Ser Val Val Gln Ser Glu Lys Ser  
1 5 10 15

<210> 26  
<211> 15  
<212> PRT



<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of  
mut.SF162 cleavage site

<400> 26

Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser  
1 5 10 15

<210> 27

<211> 15

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of native  
cleavage site in US4

<400> 27

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Arg  
1 5 10 15

<210> 28

<211> 5

<212> PRT

<213> Human immunodeficiency virus

<220>

<223> Description of Artificial Sequence: aa of second  
cleavage site in US4

<400> 28

Gln Ala Lys Arg Arg  
1 5

<210> 29

<211> 15

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: aa of mut.US4  
cleavage site

<400> 29

Ala Pro Thr Gln Ala Lys Arg Arg Val Val Gln Arg Glu Lys Ser  
1 5 10 15

<210> 30

<211> 1419

<212> DNA

<213> Human immunodeficiency virus

<400> 30

```

gtagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctatttt gtgcatcaga tgctaaagcc tatgacacag aggtacataa tgtctgggcc 120
acacatgcct gtgtaccacac agaccctaac ccacaagaaa tagtattgga aaatgtgaca 180
gaaaaatttta acatgtggaa aaataacatg gtagaacaga tgcattgagga tataatcagt 240
ttatgggatac aaagtctaaa gccatgtgta aagttaaccc cactctgtgt tactctacat 300
tgcactaatt tgaagaatgc tactaatacc aagagtagta attggaaaga gatggacaga 360
ggagaaataa aaaattgctc tttcaaggtc accacaagca taagaaataa gatgcagaaa 420
gaatatgcac ttttttataa acttgatgta gtaccaatag ataattgataa tacaagctat 480
aaattgataa attgtaacac ctcatgcatt acacaggcct gtccaaagggt atcctttgaa 540
ccaattccca tacattattg tgccccggct ggttttgcca ttctaaagtg taatgataag 600
aagttcaatg gatcaggacc atgtacaaat gtcagcacag tacaatgtac acatggaatt 660
aggccagtag tgtcaactca attgctgtta aatggcagtc tagcagaaga aggggtagta 720
attagatctg aaaatttcac agacaatgct aaaactataa tagtacagct gaaggaaatt 780
gtagaaatta attgtacaag acctaacaat aatacaagaa aaagtataac tataggaccg 840
gggagagcat tttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900
attagtggag aaaaaatggaa taacacttta aaacagatag ttacaaaatt acaagcacia 960
tttgggaata aaacaatagt ctttaagcaa tcctcaggag gggaccacga aattgtaatt 1020
cacagtttta attgtggagg ggaatttttc tactgtaatt caacacagct ttttaattag 1080
acttggaaata atactatagg gccaaataac actaatggaa ctatcacact cccatgcaga 1140
ataaaacaaa ttataaacag gtggcaggaa gtaggaaaag caatgtatgc ccctcccatc 1200
agaggacaaa ttagatgctc atcaaatatt acaggactgc tattaacaag agatgggtgt 1260
aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatat gagggacaa 1320
tggagaagtg aattatataa atataaagta gtaaaaattg agccattagg agtagcacc 1380
accaaggcaa agagaagagt ggtgcagaga gaaaaaaga 1419

```

&lt;210&gt; 31

&lt;211&gt; 1932

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 31

```

gtagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctatttt gtgcatcaga tgctaaagcc tatgacacag aggtacataa tgtctgggcc 120
acacatgcct gtgtaccacac agaccctaac ccacaagaaa tagtattgga aaatgtgaca 180
gaaaaatttta acatgtggaa aaataacatg gtagaacaga tgcattgagga tataatcagt 240
ttatgggatac aaagtctaaa gccatgtgta aagttaaccc cactctgtgt tactctacat 300
tgcactaatt tgaagaatgc tactaatacc aagagtagta attggaaaga gatggacaga 360
ggagaaataa aaaattgctc tttcaaggtc accacaagca taagaaataa gatgcagaaa 420
gaatatgcac ttttttataa acttgatgta gtaccaatag ataattgataa tacaagctat 480
aaattgataa attgtaacac ctcatgcatt acacaggcct gtccaaagggt atcctttgaa 540
ccaattccca tacattattg tgccccggct ggttttgcca ttctaaagtg taatgataag 600
aagttcaatg gatcaggacc atgtacaaat gtcagcacag tacaatgtac acatggaatt 660
aggccagtag tgtcaactca attgctgtta aatggcagtc tagcagaaga aggggtagta 720
attagatctg aaaatttcac agacaatgct aaaactataa tagtacagct gaaggaaatt 780
gtagaaatta attgtacaag acctaacaat aatacaagaa aaagtataac tataggaccg 840
gggagagcat tttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900
attagtggag aaaaaatggaa taacacttta aaacagatag ttacaaaatt acaagcacia 960
tttgggaata aaacaatagt ctttaagcaa tcctcaggag gggaccacga aattgtaatt 1020
cacagtttta attgtggagg ggaatttttc tactgtaatt caacacagct ttttaattag 1080
acttggaaata atactatagg gccaaataac actaatggaa ctatcacact cccatgcaga 1140
ataaaacaaa ttataaacag gtggcaggaa gtaggaaaag caatgtatgc ccctcccatc 1200
agaggacaaa ttagatgctc atcaaatatt acaggactgc tattaacaag agatgggtgt 1260
aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatat gagggacaa 1320
tggagaagtg aattatataa atataaagta gtaaaaattg agccattagg agtagcacc 1380
accaaggcaa agagaagagt ggtgcagaga gaaaaaagag cagtgcagct agtagctatg 1440
ttccttgggt tcttgggagc agcaggagc actatgggag caggtgcact gacgtgacg 1500
gtacaggcca gacaattatt gtctgtgata gtgcaacagc agaacaattt gctgagagct 1560
attgaggcgc aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca 1620
agagtcttgg ctgtggaaag atacctaaag gatcaacagc tcctagggat ttggggttgc 1680

```

tctggaaaac tcatttgcac cactgctgtg ccttggaatg ctagttggag taataaatct 1740  
 ctggatcaga tttggaataa catgacctgg atggagtggg agagagaaat tgacaattac 1800  
 acaaacttaa tatacacctt aattgaagaa tcgcagaacc aacaagaaaa gaatgaacaa 1860  
 gaattattag aattggataa gtgggcaagt ttgtggaatt ggtttgacat atcaaaatgg 1920  
 ctgtggtata ta 1932

&lt;210&gt; 32

&lt;211&gt; 2457

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 32

gtagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60  
 actctatttt gtgcatcaga tgctaaagcc tatgacacag aggtacataa tgtctggggc 120  
 acacatgcct gtgtacccac agacctaac ccacaagaaa tagtattgga aaatgtgaca 180  
 gaaaatttta acatgtggaa aaataacatg gtagaacaga tgcattgaga tataatcagt 240  
 ttatgggac aaagtctaaa gccatgtgta aagttaaccc cactctgtgt tactctacat 300  
 tgcactaatt tgaagaatgc tactaatacc aagagtagta attggaaaga gatggacaga 360  
 ggagaaataa aaaattgtct tttcaaggtc accacaagca taagaaataa gatgcagaaa 420  
 gaattatgcac ttttttataa acttgatgta gtaccaatag ataatagataa tacaagctat 480  
 aaattgataa attgtaacac ctacgtcatt acacaggcct gtccaaaggc atcctttgaa 540  
 ccaattccca tacattattg tgccccggct gggttttgcga ttctaaagt taatgataag 600  
 aagttcaatg gatcaggacc atgtacaaat gtcagcacag tacaatgtac acatggaatt 660  
 aggcagtag tgtcaactca attgctgtta aatggcagtc tagcagaaga aggggtagta 720  
 attagatctg aaaatttcac agacaatgct aaaactataa tagtacagct gaaggaaatct 780  
 gtagaaatta attgtacaag acctaacat aatacaagaa aaagtataac tataggaccg 840  
 gggagagcat tttatgcaac aggagacata ataggagata taagacaagc acattgtaac 900  
 attagtggag aaaaatggaa taacacttta aaacagatag ttacaaaatt acaagcacia 960  
 tttgggaata aaacaatagt ctttaagcaa tcttcaggag gggacccaga aattgtaatg 1020  
 cacagtttta attgtggagg ggaatttttc tactgtaatt caacacagct ttttaatagt 1080  
 acttgggaata atactatagg gccaaataac actaatggaa ctatcacact cccatgcaga 1140  
 ataaaaacaa ttataaacag gtggcaggaa gtaggaaaag caatgtatgc ccctcccatc 1200  
 agaggacaaa ttagatgtct atcaaatatt acaggactgc tattaacaag agatgggtgt 1260  
 aaagagatca gtaacaccac cgagatcttc agacctggag gtggagatat gagggacaat 1320  
 tggagaagtg aattatataa atataaagta gtaaaaattg agccattagg agtagcacc 1380  
 accaaggcaa agagaagagt ggtgcagaga gaaaaaagag cagtgcagct agtagctatg 1440  
 ttccttgggt tcttgggagc agcaggaagc actatgggag caccgtcact gacgtgacg 1500  
 gtacaggcca gacaattatt gtctggtata gtgcaacagc agaacaattt gctgagagct 1560  
 attgaggcgc aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca 1620  
 agagtcctgg ctgtggaaag atacctaaag gatcaacagc tcttagggat ttgggggttg 1680  
 tctggaaaac tcatttgcac cactgctgtg ccttggaatg ctagttggag taataaatct 1740  
 ctggatcaga tttggaataa catgacctgg atggagtggg agagagaaat tgacaattac 1800  
 acaaacttaa tatacacctt aattgaagaa tcgcagaacc aacaagaaaa gaatgaacaa 1860  
 gaattattag aattggataa gtgggcaagt ttgtggaatt ggtttgacat atcaaaatgg 1920  
 ctgtggtata taaaaatatt cataatgata gtaggaggtt tagtaggttt aaggatagtt 1980  
 ttactgtgc tttctatagt gaatagagtt aggcagggat actcaccatt atcatttcag 2040  
 accgcttcc cagccccaag gggacccgac agggccgaag gaatcgaaga agaagggtgga 2100  
 gagagagaca gagacagatc cagtccatta gtgcatggat tattagcact catctgggac 2160  
 gatctacgga gcctgtgctt cttcagctac caccgcttga gagacttaat cttgattgca 2220  
 gcgaggattg tggaaacttct gggacgcagg ggggtgggaag ccctcaagta ttgggggaat 2280  
 ctctgcagat attggattca ggaactaaag aatagtgtct ttagtttgtt tgatgccata 2340  
 gctatagcag tagctgaggg gacagatagg attatagaag tagcacaag aattggtaga 2400  
 gcttttctcc acatacctag aagaataaga cagggtcttt aaagggtctt gctataa 2457

&lt;210&gt; 33

&lt;211&gt; 1453

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp120.modSF162

&lt;400&gt; 33

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcggtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgcacc aacctgaaga acgccaccaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccag 600
gcctgcccga aggtgagctt cgagcccatc cccatccact actgcgcccc cgccggttc 660
gccatctga agtgcaacga caagaagttc aacggcagcg gcccctgcac caacgtgagc 720
accgtgcagt gcaccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 780
agcctggcgg aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca ccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgc gccttctacg ccaccggcga catcatcggc 960
gacatccgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcagg ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagacccc agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
ggcaccatca cctgccccct cgcacatcaag cagatcatca accgctggca ggaggtgggc 1260
aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga ccgcgcagcg cggcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcgggc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tgg                                     1453

```

&lt;210&gt; 34

&lt;211&gt; 1387

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp120.modSF162.delV2

&lt;400&gt; 34

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcggtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgcacc aacctgaaga acgccaccaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggcgccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccaggg cctgccccaa ggtgagcttc 540
gagcccatcc ccattcacta ctgcgcccc cgccgcttcg ccattcctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tgggtgagcac ccagctgctg ctgaacggca gcctggccga ggagggcggtg 720
gtgatccgca gcgagaactt caccgacaa gcgaagacca tcactgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaa aacaacaccc gcaagagcat caccatcggc 840
cccggccgcg ccttctacgc accggcgag atcatcggg acatccgcca ggccccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcgcca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagacccca gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgccctgc 1140

```

```

cgcacaaagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cgcacacccc 1200
atccgcggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaaacac caccgagatc ttccgccccg gcggcgccga catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccccct gggcggtggc 1380
ccacca
1387

```

&lt;210&gt; 35

&lt;211&gt; 1323

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp120.modSF162.delV1V2

&lt;400&gt; 35

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttccgcccag cgcctgggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggccctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgctg tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtggggcg cggcaactgc cagaccagcg tgatcaccca ggccctgcccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcgggctt cgccatcctg 480
aagtgcacag acaagaagtt caacggcagc ggccctgca ccaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtgggtgag acccagctgc tgctgaacgg cagcctggcc 600
gaggagggcg tgggtgatccg cagcgagaac ttaccgaca acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc acccgcccc acaacaacac ccgcaagagc 720
atcaccatcg gccccggccg cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggcccaact gcaacatcag cggcgagaag tggaaacaac cctgaagca gatcgtgacc 840
aagctgcagg cccagttcgg caacaagacc atcgtgttca agcagagcag cggcgggcgac 900
cccgagatcg tgatgcacag cttcaactgc ggcggcgagt tcttctactg caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcggcccc acaacaccaa cggcaccatc 1020
accctgccct gccgcatcaa gcagatcatc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccatccgccc ccagatcccg tgcagcagca acatcacccg cctgctgctg 1140
acccgcgacg gcggcaagga gatcagcaac accaccgaga tcttccgccc cggcgggcgcc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctgggctgtg ccccccacaa ggccaagcgc cgcgtggtgc agcgcgagaa gcgctaactc 1320
gag
1323

```

&lt;210&gt; 36

&lt;211&gt; 2025

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp140.modSF162

&lt;400&gt; 36

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttccgcccag cgcctgggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggccctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgctg tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgcac aacctgaaga acgccaacaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcacccag 600

```

```

gcctgccccca aggtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 660
gccatcctga agtgcaacga caagaagttc aacggcagcg gcccctgcac caacgtgagc 720
accgtgcagt gcacccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 780
agcctggccg aggggggctt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgggcgc gccttctacg ccaccggcga catcatcggc 960
gacatccgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagcaccg agctgttcaa cagcacctgg aacaacacca tcggcccaa caacaccaac 1200
ggcaccatca cctgcccctg ccgcatcaag cagatcatca accgctggca ggaggtggg 1260
aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga cccgcgacgg cggcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcgggc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tgggcgtggc cccaccaag gccaaagcgc gcgtggtgca gcgcgagaag 1500
cgcgcctgta cctggggcgc catgttccct ggcttccctg gcgcgcggc cagcaccatg 1560
ggcgcgcgca gcctgaccct gaccgtgcag gcccgccagc tgctgagcgg catcgtgcag 1620
cagcagaaca acctgctgca cgccatcagc gccacgagc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcgcgctg ctggccgtgg agcgctacct gaaggaccag 1740
cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactgggtcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

```

&lt;210&gt; 37

&lt;211&gt; 1944

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.modSF162.delV2

&lt;400&gt; 37

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
ccgctgtgga aggagggcac caccacctg ttctgcgcca gcgacgcaa ggctacgac 180
accgaggtgc acaacgtgtg gggcaccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gcaactgcaa aacctgaaga acgccaccaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgggcgac 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcccc ggcggcttcg ccctcctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tgggtgagcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagacca tcctcgtgca gctgaaggag 780
agcgtggaga tcaactgcac cgcgcccaac aacaacaccc gcaagagcat caccatcggc 840
cccgcccgcg ctttctacgc caccggcgac atcatcggcg acatccgcca ggccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaa gacagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccac gctgttcaac 1080
agcacctgga acaacaccat cgcccccaac aacaccaacg gcaccatcac cctgccctgc 1140
cgcacatcagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cgccccccc 1200
atccgcggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccc ggcggcgga catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcgtggcc 1380
cccaccaagg ccaagcgcgg cgtggtgcag cgcgagaagc gcgcgctgac cctgggcgcc 1440

```

```

atgttcctgg gcttcctggg cgccgcccgc agcaccatgg gcgcccgcag cctgaccctg 1500
accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgctgc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggcc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatctaact cgag 1944

```

&lt;210&gt; 38

&lt;211&gt; 1944

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.modSF162.delV1/V2

&lt;400&gt; 38

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcca ggccctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgctg tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccgcgccgag atcaagaact gcagcttcaa ggtgggccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagc cctgccccaa ggtgagcttc 540
gagcccatcc ccattcacta ctgcgcccc gccggtctcg ccattctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagacca tcatcgtaga gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacacc gcaagagcat caccatcggc 840
cccgcccgcg ccttctacgc caccggcgac atcatcggcg acatccgccâ ggccccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgacca gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccâ gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgccctgc 1140
cgcacaaagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cggccccccc 1200
atccgcggcc agatccgctg cagcagcaac atcaccggc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccc gcggcgccga catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcggtggc 1380
cccaccaagg ccaagcgccg cgtgtgtgag cgcgagaagc gcgccgtgac cctggggcgc 1440
atgttcctgg gcttcctggg cgccgcccgc agcaccatgg gcgcccgcag cctgaccctg 1500
accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgctgc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggcc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatctaact cgag 1944

```

&lt;210&gt; 39

&lt;211&gt; 2025

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut.modSF162

&lt;400&gt; 39

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcca ggcttacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgctgct ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gactgcacc aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gcctgttctt acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccagg 600
gcctgcccca aggtgagcct cgagcccatc cccatccact actgcgcccc cgccggcttc 660
gccatcctga agtgaacga caagaagttc aacggcagcg gccctgcac caacgtgagc 720
accgtgcagt gacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggcgg aggagggcgt ggtgatccgc agcgagaact taccgacaa cgccaagacc 840
atcatcgtgc agtgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg cccggccgc gccttctacg ccaccggcga catcatcggc 960
gacatccgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagcaccg agctgttcaa cagcacctgg aacaacacca tcggcccca caacaccaac 1200
ggcaccatca ccctgccctg ccgcatacag cagatcatca accgtggca ggaggtgggc 1260
aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga ccgcgcagcg cggaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcgggc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tgggcgtggc cccaccaag gccaaagccc gcgtggtgca gcgcgagaag 1500
agcgccgtga ccctgggcgc catgttcctg ggcttcctgg gcgcgcggc cagcaccatg 1560
ggcggccgca gcctgaccct gaccgtgcag gcccgccagc tgetgagcgg catcgtgcag 1620
cagcagaaca acctgctgcg cgccatcgag gccagcagc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcccgcgct ctggccgtgg agcgctacct gaaggaccag 1740
cagctgctgg gcactgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggag cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

```

&lt;210&gt; 40

&lt;211&gt; 1944

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut.modSF162.delV2

&lt;400&gt; 40

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcca ggcttacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgctgct ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gactgcacc aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtggcgccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540

```



```

gagcccatcc ccatccacta ctgcgcccc ggcggcttcg ccatacctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggcga ggaggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaacca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacacc gcaagagcat caccatcggc 840
cccgcccgcg ccttctacgc caccggcgac atcatcggcg acatccgcca ggcactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tctgaccaa gctgcaggcc 960
cagttcgga acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcacca gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcacatcac cctgccctgc 1140
cgcatcaagc agatcatcaa ccgctggcag gaggtgggca aggcattgta cgcaccccc 1200
atccgcggcc agatccgctg cagcagcaac atcaccggc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccc gcggcgcgca catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcgtggcc 1380
cccaccaagg ccaagcgccg cgtggtgtag cgcgagaaga gcgcgtgac cctgggcgcc 1440
atgttccctg gcttccctgg cgccgcccgc agcaccatgg gcgcgcgag cctgacctg 1500
accgtgcagg ccgcccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgctg tggcgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgacgcggca agctgatctg caccaccgcc gtgccctgga acccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcg gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggccc agcctgtgga actggttcga catcagcaag 1920
tggtgtggt acatctaact cgag

```

1944

&lt;210&gt; 41

&lt;211&gt; 1836

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.mut.modSF162.delV1/V2

&lt;400&gt; 41

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgcgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
ccgctgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaactgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgggcgc cggcaactgc cagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcccgtt cggcatcctg 480
aagtgaacg acaagaagtt caacggcagc ggcccctgca ccaactgag caccgtgcag 540
tgcaccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggaggcgg tggtagccg cagcgagaac ttaccgaca acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gcccggccg cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggccact gcaacatcag cggcgagaag tggaaacaac ccctgaagca gatcgtgacc 840
aagctgcagg ccagttcgg caacaagacc atcgtgttca agcagagcag cggcgggcag 900
ccgagatcg tgatgcacag cttcaactgc ggcgcgagtg tcttctactg caacagcacc 960
cagctgttca acagcactg gaacaacacc atcgccccca acaacaccaa cggcaccatc 1020
accctgccct gccgcatcaa gcagatcatc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccatccgagg ccagatccgc tgcagcagca acatcaccgg cctgctgctg 1140
accgcgacg gcggcaagga gatcagcaac accaccgaga tcttccgccc cggcgggcgc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctggcgctgg cccccacaa ggccaagcgc cgcgtggtgc agcgcgagaa gagcgccgtg 1320
acctggggcg ccatgttcc tgggttccct ggcgcgcgg gcagcaccat gggcgcccgc 1380
agcctgaccc tgaccgtgca ggcgcgcag ctgctgagcg gcacgtgca gcagcagaac 1440

```

```

aacctgctgc ggcgccatcga gggccagcag cacctgctgc agctgaccgt gtggggcatc 1500
aagcagctgc agggcccgct gctggccgtg gagcgctacc tgaaggacca gcagctgctg 1560
ggcatctggg gctgcagcgg caagctgac tgcaccaccg ccgtgccctg gaacgccagc 1620
tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
gacatcagca agtggtgtg gtacatctaa ctcgag 1836

```

&lt;210&gt; 42

&lt;211&gt; 2025

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.mut7.modSF162

&lt;400&gt; 42

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggagggcac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgctg tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccaccaa caccaagagc 420
agcaactgga aggagatgga cgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccag 600
gcctgcccc aagtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 660
gccatcctga agtgcaacga caagaagttc aacggcagcg gccctgcac caacgtgagc 720
accgtccagt gcacccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 780
agcctggcgg aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgctg agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgggcgc gccttctacg ccaccggcga catcatcggc 960
gacatccgct agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtagcca agctgcaggc ccagtctggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagcaccc agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
ggcaccatca ccctgccctg ccgcatcaag cagatcatca accgctggca ggaggtgggc 1260
aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga cccgcgacgg cggaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcgggc acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tgggctgtgg ccccaacca gcatcagca gcgtggtgca gagcgagaag 1500
agcggcgtga ccctggggcg catgttctct ggcttctctg gcgcccggc cagcaccatg 1560
ggcggcgcca gcctgacct gaccgtgcag gccgcccagc tgcgtgagcg catcgtgcag 1620
cagcagaaca acctgctgcg cgccatcgag gcccagcagc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcccgctg ctggccgtgg agcgctacct gaaggaccag 1740
cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccaagt ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagcggc agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

```

&lt;210&gt; 43

&lt;211&gt; 1944

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut7.modSF162.delV2

<400> 43

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccaccaa caccaagatg 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgggcgcc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540
gagccccatc ccatccacta ctgcgcccc gccggcttcg ccatacctgaa gtgcaacgac 600
aagaagttca agggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagacac ccagctgctg ctgaacggca gcctggccga ggaggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagacca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaa aacaacacc gcaagagcat caccatcggc 840
cccgcccgcg ccttctacgc caccggcgac atcatcgcg acatccgcca ggccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccca gctgttcaac 1080
agcacctgga acaacaccat cgccccaa aacaccaacg gcaccatcac cctgccctgc 1140
cgcatcaagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cgcaccccc 1200
atccgcgccc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaaac caccgagatc ttccgcccc gcggcgcgca catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcgtggcc 1380
cccaccaagg ccatcagcag cgtggtgcag agcgagaaga gcgccgtgac cctgggcgcc 1440
atgttcctgg gcttcctggg cgcgcggcg agcaccatgg gcgcgcgag cctgaccctg 1500
accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg ccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgtgc tggccgtgga gcgtacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggc agcctgtgga actggttcga catcagcaag 1920
tggtgtggt acatctaact cgag 1944

```

<210> 44

<211> 1836

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
gp140.mut7.modSF162.delV1/V2

<400> 44

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgggcgc cggcaactgc cagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgcc ccgcggctt cgccatcctg 480
aagtgaacg acaagaagtt caacggcagc ggccctgca ccaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggagggcg tggatgatcc cagcgagaac ttcaccgaca acgccaagac catcatcgtg 660

```

```

cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gccccggccg cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggcccaact gcaacatcag cggcgagaag tggaaacaaca ccctgaagca gatcgtgacc 840
aagctgcagg ccaggttcgg caacaagacc atcgtgttca agcagagcag cggcggcgac 900
cccgagatcg tgatgcacag cttcaactgc ggcggcgagt tcttctactg caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcggcccca acaacaccaa cggcaccatc 1020
accctgcccc gccgcatcaa gcagatcatc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccatccggcg ccagatccgc tgcagcagca acatcacccg cctgctgctg 1140
acccgcgacg gcggcaagga gatcagcaac accaccgaga tcttccgccc cggcggcggc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctgggcgtgg cccccaccaa ggccatcagc agcgtggtgc agagcgagaa gagcgccgtg 1320
accctggggc ccatgttccct gggcttccctg ggcgcggccg gcagcaccat gggcgccccg 1380
agcctgaccc tgaccgtgca ggcccggcag ctgctgagcg gcacgtgca gcagcagaac 1440
aacctgctgc gcgccatcga ggcccagcag cacctgctgc agctgaccgt gtggggcatc 1500
aagcagctgc agggcccgct gctggccgtg gagcgctacc tgaaggacca gcagctgctg 1560
ggcatctggg gctgcagcgg caagctgac tgcaccaccg ccgtgccctg gaacgccagc 1620
tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
gacatcagca agtggctgtg gtacatctaa ctcgag 1836

```

&lt;210&gt; 45

&lt;211&gt; 2025

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.mut8.modSF162

&lt;400&gt; 45

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgcctgtggg aagctgtggg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca_ ggctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gcaactgcac aacctgaaga acgccaccaa caccaagagc 420
agcaactgga agggatgga cggcgcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtgggtgcc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccag 600
gcctgcccc aagtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 660
gccatccctga agtgcaacga caagaagttc aacggcagcg gccctgcac caacgtgagc 720
accgtgcagt gcacccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 780
agcctggcgg agggggcggt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgggcgc gccttctacg ccaccggcga catcatcggc 960
gacatccgcc agggcccatc caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggg ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagcacc agctgttcaa cagcacctgg aacaacacca tcggccccaa caacaccaac 1200
ggcaccatca ccctgccctg ccgcatcaag cagatcatca accgctggca ggaggtggg 1260
aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcacggc 1320
ctgctgctga cccgcgacgg cggcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcggcg acatgcgca caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tgggcgtggc cccaccatc gccatcagca gcgtggtgca gagcgagaag 1500
agcgccgtga ccctgggcgc catgttccctg ggcttccctg gcgcgcggcg cagcaccatg 1560
ggcgcccgca gcctgacctt gaccgtgcag gcccggcagc tgctgagcgg catcgtgcag 1620
cagcagaaca acctgctgcg cgccatcagc gcccagcagc acctgctgca gctgaccgtg 1680

```

```

tggggcatca agcagctgca gggccgctg ctggccgtgg agcgtacct gaaggaccag 1740
cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggctgtgg tacatctaac tcgag 2025

```

&lt;210&gt; 46

&lt;211&gt; 1944

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut8.modSF162.delV2

&lt;400&gt; 46

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgcgctggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgctgct ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgcacc aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccgcgcgag atcaagaact gcagcttcaa ggtgggcgcc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagc cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcccc gccggttcg ccatectgaa gtgcaacgac 600
aagaagttca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaagacca tcactgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac gcaagagcat caccatcggc 840
cccgcccgcg ccttctacgc caccggcgac atcatcgcg acatccgcca ggccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgacca gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgacc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccac gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgccctgc 1140
cgcatcaagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cgccccccc 1200
atccgcgggc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgagcggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccc gcgggcgga catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcgtggcc 1380
cccaccatcg ccatcagcag cgtggtgca agcgagaaga gcgccgtgac cctgggcgcc 1440
atgttcctgg gcttcctggg cgccgcccgc agcaccatgg gcgcccgag cctgaccctg 1500
accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg ccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
ggcgcgctgc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgga gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggccc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatctaact cgag 1944

```

&lt;210&gt; 47

&lt;211&gt; 1836

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut8.modSF162.delV1/V2

&lt;400&gt; 47

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgctgct ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaaca catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtggggcg cggaactgc cagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat cccatccac tactgcgccc ccgcccgtt cgcctcctg 480
aagtgaacg acaagaagt caacggcagc gggccctgca ccaacgtgag caccgtgacg 540
tgacccacg gcattccgcc cgtggtgagc acccagctgc tgcgaacgg cagcctggcc 600
gaggagggcg ggtggtacg cagcgagaac ttaccgaca acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gccccggccg cgcttctac gccaccggcg acatcatcgg cgacatccgc 780
caggccact gcaacatcag cggcgagaag tggacaaca ccctgaagca gatcgtgacc 840
aagctgcagg ccaggttcgg caacaagacc atcgtgttca agcagagcag cggcgcgac 900
cccgagatcg tgatgcacg ctccaactgc ggcggcgagt tcttctactg caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcgcccca acaacaccaa cggcaccatc 1020
accctgccct gccgcacaa gcagatcatc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccatccgccc ccagatccgc tgcagcagca acatcacccg cctgctgctg 1140
accgcgacg gcggcaagga gatcagcaac accaccgaga tcttcgccc cggcgcgccg 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcgagccc 1260
ctgggctggt cccccaccat cgccatcagc agcgtgtgct agagcgagaa gagcgccgtg 1320
accctgggcg ccatgttctt gggcttctct ggcgcgccc gcagcaccat gggcgccccg 1380
agcctgaccc tgaccgtgca gggccgccc ctgctgagcg gcacgtgca gcagcagaac 1440
aacctgctgc gcgccatcga gggccagcag cacctgctgc agctgaccgt gtggggcatc 1500
aagcagctgc aggcccgctg gctggccgtg gacgctacc tgaaggacca gcagctgctg 1560
ggcatctggg gctgcagcgg caagctgate tgaccaccg ccgtgcccctg gaacgccagc 1620
tgagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680
gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactggttc 1800
gacatcagca agtggctgtg gtacatctaa ctcgag 1836

```

&lt;210&gt; 48

&lt;211&gt; 2547

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp160.modSF162

&lt;400&gt; 48

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcccga gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgctgct ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaaca catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgacc aacctgaaga acgccaacaa caccaagagc 420
agcaactgga aggagatgga ccgcccggcg atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcacccag 600
gcctgcccc aagtgagctt cgagcccatc cccatccact actgcgccc cgcggcttc 660
gccatcctga agtgcaacga caagaagttc aacggcagcg gccctgac caacgtgagc 720
accgtgcagt gacccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 780
agcctggccc aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gacgtggag atcaactgca ccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgc gccttctacg ccaccggcga catcatcggc 960
gacatccgcc aggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020

```

```

atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcggcgacc ccgagatcgt gatgcacagc ttcaactgcg gcggcgagtt cttctactgc 1140
aacagcacc cagctgttcaa cagcacctgg aacaacacca tcggcccca caacaccaac 1200
ggcaccatca ccctgcccctg ccgcatcaag cagatcatca accgctggca ggaggtgggc 1260
aaggccatgt acgccccccc catcccgggc cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga ccccgacggg cggcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcggcgggc acatgcgcga caactggcgc agcagctgtg acaagtacaa ggtggtgaag 1440
atcgagcccc tgggctgggc cccaccaag gccaaagcgc gcgtggtgca gcgcgagaag 1500
cgcgccgtga ccctggggcg catgttcctg ggcttcctgg gcgcgcggc cagcaccatg 1560
ggcgcccgca gcctgacctg gaccgtgcag gcccgcagc tgctgagcgg catcgtgcag 1620
cagcagaaca acctgctgcg gccatcgag gcccgagc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcccgctg ctggcctgg agcgtacct gaaggaccag 1740
cagctgctgg gcatctgggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggac cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaccaac ctgatctaca ccctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980
aactggttcg acatcagcaa gtggctgtgg tacatcaaga tcttcacat gatcgtgggc 2040
ggcctggtgg gcctgcgcgt cgtgttcacc gtgctgagca tcgtgaaccg cgtgcgccag 2100
ggctacagcc ccctgagctt ccagaccgcg tccccgcgccc ccgcgggccc cgaccgcccc 2160
gagggcatcg aggaggaggg cggcgagcgc gaccgcgacc gcagcagccc cctggtgcac 2220
ggcctgctgg ccctgatctg ggacgacctg cgcagcctgt gcctgttcag ctaccaccgc 2280
ctgcgcgacc tgatcctgat cgcgcggcgc atcgtggagc tgctggggcg ccgcggctgg 2340
gagggcctga agtactgggg caacctgctg cagtactgga tccaggagct gaagaacagc 2400
gccgtgagcc tggtcgagcg catcgccatc gccgtggccg agggcaccga ccgcatcatc 2460
gaggtggccc agcgcacatcg ccgcgccttc ctgcacatcc cccgcgcgat ccgccagggc 2520
ttcgagcgcg ccctgctgta actcgag 2547

```

&lt;210&gt; 49

&lt;211&gt; 2466 -

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp160.modSF162.delV2

&lt;400&gt; 49

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcca ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac ccaccgaccc caaccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcaactgcac aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccggggcgag atcaagaact gcagcttcaa ggtgggcgccc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcccc gcccgttctg ccctcctgaa gtgcaacgac 600
aagaagttca acggcagcgg ccctgcacc aacgtgagca ccgtgcagt caccacggc 660
atccgccccg tgggtgagcac ccagctgctg ctgaacggca gcctggccga ggagggcgtg 720
gtgatccgca gcgagaactt caccgacaac gccaaagacca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacaccc gcaagagcat caccatcggc 840
cccgcccgcg cttctacgc caccggcgac atcatcggcg acatccgcca ggccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagttc ttctactgca acagcaccga gctgttcaac 1080
agcacctgga acaacacccat cggccccaac aacacacacg gcaccatcac cctgccttgc 1140
cgcatcaagc agatcatcaa ccgctggcag gaggtgggca agggcatgta cggccccccc 1200
atccgcccgc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaacac caccgagatc ttccgcccc gcggcgcgca catgcgcgac 1320

```

```

aactggcgca gcgagctgta caagtacaag gtgggtgaaga tcgagccccc gggcgtggcc 1380
cccaccaagg ccaagcgccg cgtgggtgcag cgcgagaagc gcgccgtgac cctgggcgcc 1440
atgttccttg gcttcctggg cgcgcgcggc agcaccatgg gcgcccgcag cctgaccctg 1500
accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgtgc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgcagcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctgga caacatgacc tggatggagt gggagcgcga gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtgggccc agcctgtgga actggttcga catcagcaag 1920
tggctgtggg acatcaagat cttcatcatg atcgtggggc gcctgggtggg cctgcccctc 1980
gtgttcaccg tgctgagcat cgtgaaccgc gtgcgcccag gctacagccc cctgagcttc 2040
cagaccgctt tccccgcccc ccgcggcccc gaccgccccg agggcatcga ggaggagggc 2100
ggcgagcgcg accgcgaccg cagcagcccc ctggtgcacg gcctgctggc cctgatctgg 2160
gacgacctgc gcagcctgtg cctgttcagc taccaccgcc tgcgcgacct gatcctgac 2220
gcccgcgcga tcgtggagct gctgggcccgc cgcggtggg agggcctgaa gtactggggc 2280
aacctgctgc agtactggat ccaggagctg aagaacagcg ccgtgagcct gttcgacgcc 2340
atcgccatcg ccgtggccga gggcaccgac cgcacatcgc aggtggccca gcgcacggc 2400
cgcgccttcc tgcacatccc ccgcccgcac cgcaggggct tcgagcgcgc cctgctgtaa 2460
ctcgag
2466

```

&lt;210&gt; 50

&lt;211&gt; 2358

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp160.modSF162.delV1/V2

&lt;400&gt; 50

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcc tttcgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtag tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtggggcg cggcaactgc cagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcccgtt cggccatcctg 480
aagtgaacg acaagaagtt caacggcagc ggccccctga ccaacgtgag caccgtgcag 540
tgcacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggaggggc tggatgaccc cagcgagaac ttccaccgac acgccaagac catcatcgtg 660
cagctgaagg agagcgtgga gatcaactgc acccgcccca acaacaacac ccgcaagagc 720
atcaccatcg gccccggcgg cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggcccact gcaacatcag cggcgagaag tggacaacaa ccctgaagca gatcgtgacc 840
aagctgcagg ccagtttcgg caacaagacc atcgtgttca agcagagcag cggcggcgac 900
cccgagatcg tgatgcacag cttcaactgc ggccggcagc tcttctactg caacagcacc 960
cagctgttca acagcacctg gaacaacacc atcgccccca acaacacca cggcaccatc 1020
accctgccc tccgcatcaa gcgatcatc aaccgctggc aggaggtggg caaggccatg 1080
tacgcccccc ccatccgccc ccagatccgc tgcagcagca acatcaccgg cctgctgctg 1140
accgcgcagc gcggcaagga gatcagcaac accaccgaga tcttccgccc cggcggcggc 1200
gacatgcgcg acaactggcg cagcgagctg tacaagtaca aggtggtgaa gatcagagcc 1260
ctggggcgtg cccccaccaa ggccaagcgc cgcgtggtgc agcgcgagaa gcgcgccgtg 1320
accctggggc ccatgttcct gggcttcctg ggcgcgcggc gcagcaccat gggcggccgc 1380
agcctgaccc tgaccgtgca ggcccgcagg ctgctgagcg gcatcgtgca gcagcagaac 1440
aacctgctgc gcgccatcga ggcccagcag cactgctgac agctgaccgt gtggggcatc 1500
aagcagctgc aggcgcgctg gctggccgtg gagcgtacc tgaaggacca gcagctgctg 1560
ggcatctggg gctgcagcgg caagctgatc tgcaccaccg ccgtgccctg gaacggcagc 1620
tggagcaaca agagcctgga ccagatctgg aacaacatga cctggatgga gtgggagcgc 1680

```



```

gagatcgaca actacaccaa cctgatctac accctgatcg aggagagcca gaaccagcag 1740
gagaagaacg agcaggagct gctggagctg gacaagtggg ccagcctgtg gaactgggtc 1800
gacatcagca agtggctgtg gtacatcaag atcttcatca tgatcgtggg cggcctgggtg 1860
ggcctgcgca tcgtgttcac cgtgctgagc atcgtgaacc gcgtgcgcca gggctacagc 1920
ccctcgagct tccagaccgc cttccccgcc ccccgcggcc ccgaccgccc cgagggcac 1980
gaggaggagg gcggcgagcg cgaccgcgac cgcagcagcc ccctgggtga cggcctgtgt 2040
gcctgatctt gggacgacct gcgcagcctg tgcctgttca gctaccaccg cctgcgcgac 2100
ctgatcctga tcgcgcgccg catcgtggag ctgctgggcc gccgcggtg ggaggccctg 2160
aagtactggg gcaacctgct gcagtactgg atccaggagc tgaagaacag cgccgtgagc 2220
ctgttcgacg ccatcgccat cgccgtggcc gagggcaccg accgcatcat cgagggtggc 2280
cagcgcacgt gccgcgcctt cctgcacatc ccccgccgca tccgccaggg cttcgagcgc 2340
gcctgtctgt aactcgag                                     2358

```

&lt;210&gt; 51

&lt;211&gt; 1494

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 51

```

acaacagtc tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctgtttt gtgcatcaga tgctaaagca tacaaagcag aggcacataa cgtctgggct 120
acacatgcct gtgtacccac agaccccaac ccacaggaag taaatttaac aaatgtgaca 180
gaaaatttta acatgtggaa aaataacatg gtggaacaga tgcatgagga tataatcagt 240
ttatgggata aaagcctaaa gccatgtgta aaattaaccc cactctgtgt tacttttaat 300
tgtactgata agttgacagg tagtactaat ggcacaaata gtactagtgg cactaatagt 360
actagtggca ctaatagtac tagtactaat agtactgata gttgggaaaa gatgccagaa 420
ggagaaataa aaaactgctc ttccaatata accacaagtg taagagataa agtgcagaaa 480
gaatattctc tcttctataa acttgatgta gtaccaatag ataataataa tgctagctat 540
agattgataa attgtaatac ctcaagcatt acacaagcct gtccaaaggc atcttttgaa 600
ccaattccca tacattattg tgccccggct ggttttgcca ttctaaagtg taaagataag 660
aagttcaatg gaacaggacc atgtaaaaat gtcagcacag tacaatgcac acatggaatt 720
agaccagtag tatcaactca actgctgtta aatggcagtc tagcagaaga agagatagta 780
cttagatctg aaaatttcac agacaatgct aaaaccataa tagtacagct gaatgaatct 840
gtagaaatta attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
gggagagcat tttatgcaac aggtgatata ataggagaca taagacaagc acattgtaac 960
attagtaaag caaactggac taacacttta gaacagatag ttgaaaaatt aagagaacaa 1020
tttgggaata ataaaacaat aatctttaat tcatcctcag gaggggaccc agaaattgta 1080
tttcacagtt ttaattgttg aggggaattt ttctattgta atacatcaca actattttaa 1140
agtacctgga atattactga agaggtaaat aagactaaag aaaatgacac tatcatactc 1200
ccatgcagaa taagacaaat tataaacatg tggcaagaag taggaaaagc aatgtatgcc 1260
cctcccatca gaggacaaat taaatgttca tcaaatatta cagggtgtgt attaactaga 1320
gatggtggta ctaacaataa taggacgaac gacaccgaga ccttcagacc tgggggagga 1380
aacatgaagg acaattggag aagtgaatta tataaatata aagtagtaag aattgaacca 1440
ttaggagtag caccacacca ggcaaagaga agagtgggtg aaagagagaa aaga 1494

```

&lt;210&gt; 52

&lt;211&gt; 2007

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 52

```

acaacagtc tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctgtttt gtgcatcaga tgctaaagca tacaaagcag aggcacataa cgtctgggct 120
acacatgcct gtgtacccac agaccccaac ccacaggaag taaatttaac aaatgtgaca 180
gaaaatttta acatgtggaa aaataacatg gtggaacaga tgcatgagga tataatcagt 240
ttatgggata aaagcctaaa gccatgtgta aaattaaccc cactctgtgt tacttttaat 300
tgtactgata agttgacagg tagtactaat ggcacaaata gtactagtgg cactaatagt 360
actagtggca ctaatagtac tagtactaat agtactgata gttgggaaaa gatgccagaa 420
ggagaaataa aaaactgctc ttccaatata accacaagtg taagagataa agtgcagaaa 480

```

```

gaatattctc tcttctataa acttgatgta gtaccaatag ataatgataa tgctagctat 540
agattgataa attgtaatac ctcagtcatt acacaagcct gtccaaagggt atcttttgaa 600
ccaattccca tacattattg tgccccggct ggttttgca ttctaaaagt taaagataag 660
aagttcaatg gaacaggacc atgtaaaaat gtcagcacag tacaatgcac acatggaatt 720
agaccagtag tatcaactca actgctgtta aatggcagtc tagcagaaga agagatagta 780
cttagatctg aaaatttcac agacaatgct aaaaccataa tagtacagct gaatgaatct 840
gtagaaatta attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
gggagagcat tttatgcaac aggtgatata ataggagaca taagacaagc acattgtaac 960
attagtaaaag caaactggac taacacttta gaacagatag ttgaaaaatt aagagaacaa 1020
tttggaata ataaaaaat aatctttaat tcctcctcag gaggggaccc agaaattgta 1080
tttcacagtt ttaattgtgg aggggaattt ttctattgta atacatcaca actatttaac 1140
agtacctgga atattactga agaggtaaat aagactaaag aaaatgacac tatcatactc 1200
ccatgcagaa taagacaaat tataaacatg tggcaagaag taggaaaagc aatgtatgcc 1260
cctcccatca gaggacaaat taaatgttca tcaaatatta cagggtctgt attaactaga 1320
gatggtggtg ctaacaataa taggacgaac gacaccgaga ccttcagacc tgggggagga 1380
aacatgaagg acaattggag aagtgaatta tataaatata aagtagtaag aattgaacca 1440
ttaggagtag caccacacca ggcaaagaga agagtgggtg aaagagagaa aagagcagtg 1500
ggactaggag ctttgttcat tgggttcttg ggagcagcag gaagcactat gggcgcagcg 1560
tcagtgcagc tgacggtaca ttattgtctg gtatagtgtg acagcagaac 1620
aatgtgctga gagctattga ggcgcaacag catctgttgc aactcacggt ctggggcatc 1680
aaacagctcc aggaagaat cctggctgtg gaaagatacc taaaggatca acagctccta 1740
gggatttggg gttgctctgg aaaactcatt tgcaccacta ctgtgccttg gaactctagt 1800
tggagtaata aatctctgac tgagatttgg gataatatga cctggatgga gtgggaaaga 1860
gaaattggca ctataatagc cttaatatag aatttaattg aaatagcaca aaaccagcaa 1920
gaaaagaatg aacaagaatt attggaatta gacaagtggg caagtttgtg gaattggttt 1980
gatataacaa actggctgtg gtatata 2007

```

&lt;210&gt; 53

&lt;211&gt; 2532

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 53

```

acaacagctc tgtgggtcac agtctattat ggggtacctg tgtggaaaga agcaaccacc 60
actctgtttt gtgcatcaga tgctaaagca tacaaagcag aggcacataa cgtctgggct 120
acacatgcct gtgtaccacac agacccaac ccacaggaag taaatttaac aaatgtgaca 180
gaaaatttta acatgtggaa aaataacatg gtggaacaga tgcattgagga tataatcagt 240
ttatggggtc aaagcctaaa gccatgtgta aaattaaccc cactctgtgt tactttaaat 300
tgtactgata agttgacagg tagtactaat ggcacaaata gtactagtgg cactaatagt 360
actagtggca ctaatagtac tagtactaat agtactgata gttgggaaaa gatgccagaa 420
ggagaaataa aaaactgctc tttcaatata accacaagtg taagagataa agtgccagaa 480
gaatattctc tcttctataa acttgatgta gtaccaatag ataatgataa tgctagctat 540
agattgataa attgtaatac ctcagtcatt acacaagcct gtccaaagggt atcttttgaa 600
ccaattccca tacattattg tgccccggct ggttttgca ttctaaaagt taaagataag 660
aagttcaatg gaacaggacc atgtaaaaat gtcagcacag tacaatgcac acatggaatt 720
agaccagtag tatcaactca actgctgtta aatggcagtc tagcagaaga agagatagta 780
cttagatctg aaaatttcac agacaatgct aaaaccataa tagtacagct gaatgaatct 840
gtagaaatta attgtataag acccaacaat aatacaagaa aaagtataca tataggacca 900
gggagagcat tttatgcaac aggtgatata ataggagaca taagacaagc acattgtaac 960
attagtaaaag caaactggac taacacttta gaacagatag ttgaaaaatt aagagaacaa 1020
tttggaata ataaaaaat aatctttaat tcctcctcag gaggggaccc agaaattgta 1080
tttcacagtt ttaattgtgg aggggaattt ttctattgta atacatcaca actatttaac 1140
agtacctgga atattactga agaggtaaat aagactaaag aaaatgacac tatcatactc 1200
ccatgcagaa taagacaaat tataaacatg tggcaagaag taggaaaagc aatgtatgcc 1260
cctcccatca gaggacaaat taaatgttca tcaaatatta cagggtctgt attaactaga 1320
gatggtggtg ctaacaataa taggacgaac gacaccgaga ccttcagacc tgggggagga 1380
aacatgaagg acaattggag aagtgaatta tataaatata aagtagtaag aattgaacca 1440
ttaggagtag caccacacca ggcaaagaga agagtgggtg aaagagagaa aagagcagtg 1500
ggactaggag ctttgttcat tgggttcttg ggagcagcag gaagcactat gggcgcagcg 1560

```

```

tcagtgcgc tgacgggtaca gccagacaa ttattgtctg gtatagtgc acagcagaac 1620
aatttgctga gagctattga gccgcaacag catctgttgc aactcacggg ctggggcatc 1680
aaacagctcc aggcagaagt cctggctgtg gaaagatacc taaaggatca acagctccta 1740
gggatttggg gttgctctgg aaaactcatt tgcaccacta ctgtgccttg gaactctagt 1800
tgagtaata aatctctgac tgagatttgg gataatatga cctggatgga gtgggaaaga 1860
gaaattggca attatacagg cttaatatac aatttaattg aaatagcaca aaaccagcaa 1920
gaaaagaatg aacaagaatt attggaatta gacaagtggg caagtgtgtg gaattggttt 1980
gatataacaa actggctgtg gtatataaga atattcataa tgatagtagg aggcttgata 2040
ggtttaagaa tagtttttgc tgtactttct atagtgaata gagttaggca gggataactca 2100
ccaatatcat tgcagacccg cctcccagct cagaggggac ccgacaggcc cgaaggaatc 2160
gaagaagaag gtggagagag agacagagac agatccaatc gattagtgc tggattattg 2220
gcactcatct gggacgatct gcggagcctg tgcctcttca gctaccaccg cttgagagac 2280
ttactcttga ttgtagcgag gattgtggaa cttctgggac gcagggggtg ggaagccctc 2340
aagtattggg ggaatctcct gcagtattgg agtcaggagc taaagagtag tgctgttagt 2400
ttgtttaatg ccacagcaat agcagtagct gaagggacag ataggattat agaaatagta 2460
caaagaattt ttagagctgt aattcacata cctagaagaa taagacaggg cttggagagg 2520
gctttactat aa 2532

```

&lt;210&gt; 54

&lt;211&gt; 1599

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp120.modUS4

&lt;400&gt; 54

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggcttacaag 180
gccgagggcc acaacgtgtg ggccaccac gctgctgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagacca gcggcaccaa cagcaccagc ggcaccaaca gcaccagca caacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcaccacc 540
agcgtgcgag acaaggtgca gaaggagtac agcctgttct acaagctgga cgtggtgccc 600
atcgacaacg acaacgccag ctaccgcctg atcaactgca acaccagct gatcaccag 660
gcctgccccg aggtgagctt cgagcccatc cccatccact actgcgccc cgccggttc 720
gccatcctga agtgcaagga caagaagttc aacggcaccg gccctgcaa gaacgtgagc 780
accgtgcatg gcaccacagg catccgcccc gtggtgagca ccagctgct gctgaacggc 840
agcctggccg aggaggagat cgtgctgctg tccgagaact tcaccgacaa cgccaagacc 900
atcatcgtgc agctgaacga gtccgtggag atcaactgca tccgccccaa caacaacacg 960
cgtaagagca tccacatcgg ccccgcccg gccttctacg ccaccggcga catcatcggc 1020
gacatccgcc agggccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcgtggaga agctgcgaga gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140
agcggcgggc accccgagat cgtgttccac agcttcaact gcggcgggcga gttcttctac 1200
tgcaaacacca gccagctgtt caacagcacc tggaaactca ccgaggaggt gaacaagacc 1260
aaggagaacg acaccatcat cctgcctgc cgcatccgac agatcatcaa catgtggcag 1320
gaggtgggca aggccatgta cggccccccc atccgcgggc agatcaagtg cagcagcaat 1380
attaccggcc tgctgtgac ccgacagggc ggcaccaaca acaaccgcac caacgacacc 1440
gagaccttcc gcccggcgcg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
tacaaggtg tgcgcatcga gcccctgggc gtggccccc cccaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagcgcta agatatcgga tcctctaga 1599

```

&lt;210&gt; 55

&lt;211&gt; 1350

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp120.modUS4.del 128-194

&lt;400&gt; 55

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcggtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgggggc agggaaactg gagaccagcg tgatcaccca ggccctgccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcccgtt cgccatcctg 480
aagtgaagg acaagaagtt caacggcacc ggcccctgca agaactgag caccgtgcag 540
tgcaccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggaggaga tcgtgctgcg ctccgagaa ttaccgaca acgccaagac catcatcgtg 660
cagctgaacg agtccgtgga gatcaactg atccgcccc acaacaacac gcgtaagagc 720
atccacatcg gccccggccg cgcttctac gccaccggcg acatcatcgg cgacatccgc 780
caggccact gcaacatcag caaggccaac tggaccaaca ccctcgagca gatcgtggag 840
aagctgcgcg agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcggc 900
gaccccgaga tcgtgttcca cagcttcaac tgccggcgcg agttcttcta ctgcaacacc 960
agccagctgt tcaacagcac ctggaacatc accgaggagg tgaacaagac caaggagaac 1020
gacacatca tcctgccttg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 1080
aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa tattaccggc 1140
ctgctgctga cccgcgacgg cggcaccaac aacaaccgca ccaacgacac cgagaccttc 1200
cgccccggcg gcggaacat gaaggacaac tggcgacgag agctgtacaa gtacaagggtg 1260
gtgcgcatcg agccctggg cgtggccccc acccaggcca agcgcgcggt ggtgcagcgc 1320
gagaagcgct aagatatcgg atcctctaga                                     1350

```

&lt;210&gt; 56

&lt;211&gt; 2112

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp140.modUS4

&lt;400&gt; 56

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcggtg 120
cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgaggcgag atcaagaact gcagcttcaa catcaccacc 540
agcgtgcgcg acaagggtgca gaaggagtac agcctgttct acaagctgga cgtggtgccc 600
atcgacaacg acaacgccag ctaccgctg atcaactgca acaccagcgt gatcacccag 660
gcctgccccca aggtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 720
gccatcctga agtgcaagga caagaagttc aacggcaccg gccctgcaa gaacgtgagc 780
accgtgcagt gcaccacgg catccgccc gtggtgagca ccagctgct gctgaacggc 840
agcctggccg aggaggagat cgtgctgcgc tccgagaact tcaccgacaa cgccaagacc 900
atcatcgtgc agctgaacga gtccgtggag atcaactgca tccgccccaa caacaacacg 960
cgtaagagca tccacatcgg ccccgccgc gccttctacg ccaccggcga catcatcggc 1020
gacatccgcc aggccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcgtggaga agctgcgca gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140
agcggcgggc accccgagat cgtgttcac agcttcaact gcggcgggca gttcttctac 1200
tgcaacacca gccagctgtt caacagcacc tggaaacatca ccgaggagggt gaacaagacc 1260

```

```

aaggagaacg acaccatcat cctgccctgc cgcateccgc agatcatcaa catgtggcag 1320
gaggtgggca agggcatgta cgcccccccc atccgcggcc agatcaagtg cagcagcaat 1380
attaccggcc tgctgtgtac ccgcgacggc ggaccaaca acaaccgcac caacgacacc 1440
gagaccttcc gccccggcgg cggcaacatg aaggacaact ggcgagcga gctgtacaag 1500
tacaaggtgg tgcgcatcga gccccctggc gtggccccc cccaggccaa gcgcgcgtg 1560
gtgcagcgcg agaagcgcg cgtgggcctg ggcgccctgt tcatcggtt cctgggcgcc 1620
gccgggagca ccatgggcgc cgcctccgtg accctgaccg tgcaggcccg ccagctgctg 1680
agcggcatcg tgcagcagca gaacaacctg ctgcgcgcca tcgaggccca gcagcacctg 1740
ctgcagctga ccgtgtgggg catcaagcag ctgcaggccc gcacccctggc cgtggagcgc 1800
tacctgaagg accagcagct gctgggcac tggggctgca gcggcaagct gatctgcacc 1860
accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920
atgacctgga tggagtggga gcgcgagatc ggcaactaca ccggcctgat ctacaacctg 1980
atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
tgggcccagc tgtggaactg gttcgacatc accaactggc tgtggtacat ctaagatatc 2100
ggatcctcta ga                                     2112

```

&lt;210&gt; 57

&lt;211&gt; 2112

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.mut.modUS4

&lt;400&gt; 57

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
ccctgttgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgaggcgag atcaagaact gcagcttca catcaccacc 540
agcgtgcgcg acaaggtgca gaaggagtac agcctgttct acaagctgga cgtggtgccc 600
atcgacaacg acaacgccag ctaccgcctg atcaactgca acaccagcgt gatcaccacg 660
gcctgcccc aagtgagctt cgagccatc cccatccact actgcgcccc cgccggcttc 720
gccatcctga agtgcaagga caagaagttc aacggcaccg gcccctgcaa gaacgtgagc 780
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 840
agcctggccg agggaggagat cgtgctgcgc tccgagaact tcaccgaaa cgccaagacc 900
atcatcgtgc agctgaacga gtccgtggag atcaactgca tccgccccaa caacaacacg 960
cgtaagagca tccacatcgg ccccgccgc gccttctacg ccacggcgga catcatcgcc 1020
gacatccgcc agggccactg caacatcagc aaggccaact ggaccaaac cctcgagcag 1080
atcgtggaga agctgcgcga gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140
agcggcgcg accccgagat cgtgttcac agcttcaact gcggcgcgga gttcttctac 1200
tgacaacacca gccagctgtt caacagcacc tggaacatca ccgaggaggt gaacaagacc 1260
aaggagaacg acaccatcat cctgccctgc cgcateccgc agatcatcaa catgtggcag 1320
gaggtgggca agggcatgta cgcccccccc atccgcggcc agatcaagtg cagcagcaat 1380
attaccggcc tgctgtgtac ccgcgacggc ggaccaaca acaaccgcac caacgacacc 1440
gagaccttcc gccccggcgg cggcaacatg aaggacaact ggcgagcga gctgtacaag 1500
tacaaggtgg tgcgcatcga gccccctggc gtggccccc cccaggccaa gcgcgcgtg 1560
gtgcagcgcg agaagagcgc cgtgggcctg ggcgccctgt tcatcggtt cctgggcgcc 1620
gccgggagca ccatgggcgc cgcctccgtg accctgaccg tgcaggcccg ccagctgctg 1680
agcggcatcg tgcagcagca gaacaacctg ctgcgcgcca tcgaggccca gcagcacctg 1740
ctgcagctga ccgtgtgggg catcaagcag ctgcaggccc gcacccctggc cgtggagcgc 1800
tacctgaagg accagcagct gctgggcac tggggctgca gcggcaagct gatctgcacc 1860
accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920
atgacctgga tggagtggga gcgcgagatc ggcaactaca ccggcctgat ctacaacctg 1980

```

atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040  
 tggggccagcc tgtggaactg gttcgacatc accaactggc tgtggtacat ctaagatata 2100  
 ggaatccteta ga 2112

<210> 58

<211> 2181

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: gp140TM.modUS4

<400> 58

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60  
 gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120  
 cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180  
 gccgagggccc acaacgtgtg ggccaccac gcctgcgtgc ccaccgaccc caacccccag 240  
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtgga 300  
 cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360  
 acccccctgt gctgacccct gaactgcacc gacaagctga ccggcagcac caacggcacc 420  
 aacagcacca gcggcaccaa cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480  
 gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcaccacc 540  
 agcgtgcgcg acaaggtgca gaaggagtac agcctgttct acaagctgga cgtggtgccc 600  
 atcgacaacg acaacgcccgt ctaccgcccgt atcaactgca acaccagcgt gatcaccag 660  
 gcctgcccga aggtgagctt cgagcccac cccatccact actgcgccc cgccggcttc 720  
 gccatcctga agtgcaagga caagaagttc aacggcaccc gccctgcaa gaacgtgagc 780  
 accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 840  
 agcctggcgg aggaggagat cgtgctgcgc tccgagaact tcaccgacaa cgccaagacc 900  
 atcatcgtgc agtgaacga gtcctgtggag atcaactgca tccgccccaa caacaacagc 960  
 cgtaagagca tccacatcgg ccccgcccg gccttctacg ccaccggcga catcatcggc 1020  
 gacatccgcc agggccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080  
 atcgtggaga agctgcgcga gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140  
 agcggcgccg accccgagat cgtgttccac agcttcaact gcggcgccga gttcttctac 1200  
 tgcaaacacca gcaagctgtt caacagcacc tggaacatca ccgaggaggt gaacaagacc 1260  
 aaggagaacg acaccatcat cctgcccctg cgcattccgc agatcatcaa catgtggcag 1320  
 gaggtgggca agggcatgta cgcccccccc atccgcgcc agatcaagtg cagcagcaat 1380  
 attaccggcc tgcgtgctgac ccgcgacggc ggcaccaaca acaaccgcac caacgacacc 1440  
 gagaccttc gcccggcgcg cggaacatg aaggacaact ggcgcagcga gctgtacaag 1500  
 tacaaggtgg tgcgcatcga gccctgggc gtggccccca cccaggccaa gcgcccgctg 1560  
 gtgcagcgcg agaagcgcg cgtgggcctg ggcgcctgt tcatcggtt cctggcgcg 1620  
 gccgggagca ccatggcgcg cgctccgtg accctgaccg tgcaggcccg ccagctgctg 1680  
 agcggcatcg tgcagcagca gaacaacctg ctgcgcgcca tcgaggccca gcagcacctg 1740  
 ctgcagctga ccgtgtggg catcaagcag ctgcaggccc gcatcctggc cgtggagcgc 1800  
 tacctgaagg accagcagct gctgggcac tggggctgca gcggcaagct gatctgcacc 1860  
 accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920  
 atgacctgga tggagtggga gcgcgagatc ggcaactaca ccggcctgat ctacaacctg 1980  
 atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040  
 tgggcccagcc tgtggaactg gttcgacatc accaactggc tgtggtacat ccgcatcttc 2100  
 atcatgatcg tggcgccgt cgcgcctg cgcacgtgt tcgccgtgct gagcatcgtg 2160  
 taagatatcg gatcctctag a 2181

<210> 59

<211> 1818

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:  
 gp140.modUS4.delV1/V2

&lt;400&gt; 59

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca ggcacgcca ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgctgct ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgggccc 360
ggccaggcct gcccgaaggt gagcttcgag cccatccccca tccactactg cggccccgcc 420
ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
gtgagcaccg tgcagtgcac ccacggcatc cggcccgctg tgagcaccga gctgctgctg 540
aacggcagcc tggccgagga ggagatcgtg ctgctgctccg agaacttcac cgacaacgcc 600
aagaccatca tcgtgcagct gaacgagtc gtggagatca actgcatccg ccccaacaac 660
aacacgcgta agagcatcca catcgccccc ggccgcccct tctacgccac cggcgacatc 720
atcggcgaca tccgccaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccgcaaca acaagaccat catcttcaac 840
agcagcagcg gcggcgaccc cgagatcgtg ttccacagct tcaactgcgg cggcgagttc 900
ttctactgca acaccagcca gctgttcaac agcacctgga acatcaccga ggaggtgaac 960
aagaccaagg agaacgacac catcatcctg ccctgccgca tccgccagat catcaacatg 1020
tggcaggagg tgggcaaggc catgtacgcc ccccccattc gcggccagat caagtgcagc 1080
agcaatatta ccggcctgct gctgaccgcg gacggcggca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcgcccc cgggcgccgc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgcg catcgagccc ctgggcgtgg ccccaacca ggccaagcgc 1260
cgcggtggtg agcgcgagaa gcgcgcctg ggccctggcg ccctgttcat cggcttctg 1320
ggcgccgcgg ggagcaccat gggcgcgcc tccgtgacct tgaccgtgca ggcccgccag 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgctgc gcgccatcga ggcccagcag 1440
cacctgctgc agtgaccgt gtggggcatc aagcagctgc agggccgcct cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgctg ggcattctgg gctgcagcgg caagctgac 1560
tgaccacca ccgtgccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcggca actacaccgg cctgatctac 1680
aacctgatcg agatcgccca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatctaa 1800
gatatcggat cctctaga

```

&lt;210&gt; 60

&lt;211&gt; 2031

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.modUS4.delV2

&lt;400&gt; 60

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca ggcacgcca ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcctgctgct ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcggcgcc 540
ggccgcctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 600
gagcccatcc ccatccacta ctgcgcccc gcggcttcg ccactctgaa gtgcaaggac 660
aagaagttca acggcaccgg cccctgcaag aacgtgagca cgtgacgtg caccacggc 720
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggccga ggaggagatc 780
gtgctgcgct ccggaactt cccgacaac cccaagacca tcatcgtgca gctgaacgag 840
tccgtggaga tcaactgcat ccgccccaac aacaacacgc gtaagagcat ccacatcgcc 900
cccgcccgcg ccttctacgc caccggcgac atcatcgccg acatccgcca ggcccactgc 960

```

```

aacatcagca aggccaaactg gaccaacacc ctcgagcaga tctgtggagaa gctgcgcgag 1020
cagttcggca acaacaagac catcatcttc aacagcagca gcggcggcga ccccgagatc 1080
gtgttccaca gcttcaactg cgcgggcgag ttcttctact gcaacaccag ccagctgttc 1140
aacagcacct ggaacatcac cgaggagggtg aacaagacca aggagaacga caccatcatc 1200
ctgccctgcc gcatccgccca gatcatcaac atgtggcagg aggtgggcaa ggccatgtac 1260
gcccccccca tccgcgggcca gatcaagtgc agcagcaata ttaccggcct gctgctgacc 1320
cgcgacggcg gcaccaacaa caaccgcacc aacgacaccg agaccttccg ccccgcgggc 1380
ggcaacatga aggacaactg gcgcagcgag ctgtacaagt acaagggtgtt gcgcacgag 1440
cccctggggc tggccccac ccaggccaag cgccgcgtgg tgcagcgcg gaagcgcgcc 1500
gtgggccttg gcgcctgtt catcggett cctggcgccg ccgggagcac catggggcgc 1560
gcctccgtga ccctgaccgt gcaggcccgc cagctgctga gcggcatcgt gcagcagcag 1620
aacaacctgc tggcgcccat cgaggcccag cagacctgc tgcagctgac cgtgtggggc 1680
atcaagcagc tgcaggcccc catcctggcc gtggagcgct acctgaagga ccagcagctg 1740
ctgggcatct ggggctgcag cggaagctg atctgcacca ccaccgtgcc ctggaacagc 1800
agctggagca acaagagcct gaccgagatc tgggacaaca tgacctggat ggagtgggag 1860
cgcgagatcg gcaactacac cggcctgata tacaacctga tcgagatcgc ccagaaccag 1920
caggagaaga acgagcagga gctgctggag ctggacaagt gggccagcct gtggaactgg 1980
ttcgacatca ccaactggct gtggtacatc taagatatcg gatcctctag a 2031

```

&lt;210&gt; 61

&lt;211&gt; 1818

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp140.mut.modUS4.delV1/V2

&lt;400&gt; 61

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga agggggccac caccacctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gctgctgctg ccaccgacct caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtggggcgc 360
ggccaggcct gccccaaggt gagcttcgag cccatcccca tccactactg cgcccccgcc 420
ggcttcgcca tccgtgaagt caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
gtgagcaccg tgcagtgcac ccacggcatc cgccccgtgg tgagcaccga gctgctgctg 540
aacggcagcc tggccgagga ggagatcgtg ctgcgctccg agaacttcac cgacaacgcc 600
aagaccatca tctgtcagct gaacgagtc gtggagatca actgcatccg cccaacaac 660
aacacgcgta agagcatcca catcgcccc ggccgcgcct tctacgccac cggcgacatc 720
atcgccgaca tccgccaggc ccaactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccggcaaca acaagacct catcttcaac 840
agcagcagcg gcggcgaccc cgagatcgtg ttccacagct tcaactgcgg cggcgagtgc 900
ttctactgca acaccagcca gctgttcaac agcacctgga acatcaccga ggaggtgaac 960
aagaccaagg agaacgacac catcatcctg ccctgccgca tccgccagat catcaacatg 1020
tggcaggagg tgggcaaggc catgtacgcc ccccccattc gcggccagat caagtgcagc 1080
agcaatatta ccggcctgct gctgacccgc gacggcgcca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcgcccc cggcgggcggc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgcg catcgagccc ctgggcgtgg cccccacca ggccaagcgc 1260
cgcggtggtg agcgcgagaa gagcgccgtg ggcctggggc ccctgttcat cggcttcctg 1320
ggcgccgccc ggagcaccat gggcgccgcc tccgtgacct tgaccgtgca ggcccgccag 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgctgc gcgccatcga ggcccagcag 1440
cacctgctgc agctgaccgt gtggggcatc aagcagctgc agggccgcat cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgctg ggcattctgg gctgcagcgg caagctgac 1560
tgcaccacca ccgtgccttg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcggca actacaccgg cctgatctac 1680
aacctgatcg agatcgcccc gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatctaa 1800

```



gatatcggat cctctaga

1818

&lt;210&gt; 62

&lt;211&gt; 1818

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.modUS4.del 128-194

&lt;400&gt; 62

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgagggccc acaacgtgtg ggccaccacc gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgggcgcc 360
ggccaggcct gccccaaagg gagcttcgag cccatcccca tccactactg cgcccccgcc 420
ggcttcgcca tcttgaagtg caaggacaag aagtccaacg gcaccggccc ctgcaagaac 480
gtgagcaccg tgcagtgcac ccacggcatc cgccccgtgg tgagcaccga gctgctgctg 540
aacggcagcc tggcgagga ggagatcgtg ctgcgtccg agaacttcac cgacaacgcc 600
aagaccatca tctgtcagct gaacgagtc gtggagatca actgcatccg cccaacaac 660
aacacgcgta agagcatcca catcgcccc ggccgcgct tctacgccac cggcgacatc 720
atcggcgaca tccgccaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttcggaaca acaagaccat catcttcaac 840
agcagcagcg gcggcgacc cgagatcgtg ttccacagct tcaactgcgg cggcgagttc 900
ttctactgca acaccagcca gctgttcaac agcacctgga acatcaccga ggaggtgaac 960
aagaccaagg agaacgacac catcatcctg cctgcccga tccgccagat catcaacatg 1020
tggcaggagg tgggcaaggc catgtacgcc ccccccatcc gcggccagat caagtgcagc 1080
agcaatatta ccggcctgct gctgacccgc gacggcgcca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcgccc cgccggcgcc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgcg catcgagccc ctggcgctgg cccccacca ggccaagcgc 1260
cgcgtggtgc agcgcgagaa gagcgccgtg ggccgtggcg cctgttcat cggcttctg 1320
ggcgccgccc ggagcaccat gggcgccgcc tccgtgaccc tgaccgtgca ggcccgccag 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgctgc gcgccatcga ggcccagcag 1440
cacctgctgc agctgaccgt gtggggcatc aagcagctgc agggccgcat cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgctg ggcattctgg gctgcagcgg caagctgatc 1560
tgcaccacca ccgtgccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcgga actacaccg cctgatctac 1680
aacctgatcg agatcgcccc gaaccagcag gagaagaac agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatctaa 1800
gatatcggat cctctaga 1818

```

&lt;210&gt; 63

&lt;211&gt; 1863

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp140.mut.modUS4.del 128-194

&lt;400&gt; 63

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgagggccc acaacgtgtg ggccaccacc gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300

```

```

cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgggggc aggggaactgc gagaccagcg tgatcaccca ggcctgcccc 420
aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcccggctt cgccatcctg 480
aagtgcagg acaagaagtt caacggcacc ggcccctgca agaacgtgag caccgtgcag 540
tgacccacg gcatccgccc cgtggtgagc acccagctgc tgctgaacgg cagcctggcc 600
gaggaggaga tcgtgctgcg ctccgagAAC ttcaccgaca acgccaagac catcatcgtg 660
cagctgaacg agtccgtgga gatcaactgc atccgcccc acaacaacac gcgtaagagc 720
atccacatcg gccccggccg cgccttctac gccaccggcg acatcatcgg cgacatccgc 780
caggccact gcaacatcag caaggccaac tggaccaaca ccctcgagca gatcgtggag 840
aagctgcgcg agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcggc 900
gaccccgaga tcgtgttcca cagcttcaac tgcggcggcg agttcttcta ctgcaacacc 960
agccagctgt tcaacagcac ctggaacatc accgaggagg tgaacaagac caaggagaac 1020
gacacccatca tcctgcccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 1080
aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa tattaccggc 1140
ctgctgctga ccgcgacgg cggcaccaac aacaaccgca ccaacgacac cgagaccttc 1200
cgccccggcg gcggcaacat gaaggacaac tggcgacgg agctgtacaa gtacaagggtg 1260
gtgcgcatcg agcccctggg cgtggccccc acccaggcca agcgcgcgct ggtgcagcgc 1320
gagaagagcg ccgtgggccc ttcacggct tcctgggcgc cgccgggagc 1380
accatgggcg ccgctccgt gaccctgacc gtgcaggccc gccagctgct gagcggcatc 1440
gtgcagcagc agaacaacct gctgcgcgcc atcgaggccc agcagcacct gctgcagctg 1500
accgtgtggg gcatcaagca gctgcaggcc cgcacccctg ccgtggagcg ctacctgaag 1560
gaccagcagc tgctgggcat ctggggctgc agcggcaagc tgatctgcac caccaccgtg 1620
ccctggaaca ccaagagc ctgaccgaga tctgggacaa catgacctgg 1680
atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct gatcgagatc 1740
gccagaacc agcaggagaa gaacgagcag gagctgctgg agctggacaa gtgggccagc 1800
ctgtggaact ggttcgacat caccaactgg ctgtggtaca tctaagatat cggatcctct 1860
aga

```

1863

&lt;210&gt; 64

&lt;211&gt; 2634

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: gp160.modUS4

&lt;400&gt; 64

```

gaattcgcca ccattggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccacc gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccaa cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgaggggcag atcaagaact gcagcttcaa catcaccacc 540
agcgtgcgcg acaagtgca gaaggagtac agcctgttct acaagctgga cgtggtgccc 600
atcgacaacg acaacgccag ctaccgcctg atcaactgca acaccagcgt gatcaccag 660
gcctgcccc aagtgagctt cgagcccatc cccatccact actgcgcccc cgccggcttc 720
gccatcttga agtgcaagga caagaagttc aacggcaccg gcccctgcaa gaacgtgagc 780
accgtgcagt gcacccacgg catccgcccc gtggtgagca ccagctgct gctgaacggc 840
agcctggccg aggaggagat cgtgctgcgc tccgagaact tcaccgaaa cgccaagacc 900
atcatcgtgc agctgaacga gtccgtggag atcaactgca tccgccccaa caacaacacg 960
cgtaagagca tccacatcgg ccccgcccg gccttctacg ccaccggcga catcatcggc 1020
gacatccgcc agggccactg caacatcagc aaggccaact ggaccaacac cctcgagcag 1080
atcgtggaga agctgcgcga gcagttcggc aacaacaaga ccatcatctt caacagcagc 1140
agcggcggcg accccgagat cgtgttccac agcttcaact gcggcgcgga gttcttctac 1200
tgcaacacca gccagctggt caacagcacc tggaaatca ccgaggaggt gaacaagacc 1260
aaggagaacg acaccatcat cctgcctgca cgcattccgc agatcatcaa catgtggcag 1320

```

```

gagggtgggca agggcatgta cggccccccc atccgcggcc agatcaagtg cagcagcaat 1380
attaccggcc tgetgtgtgac cgcgcacggc ggcaccaaca acaaccgcac caacgacacc 1440
gagaccttcc gccccggcgg cggcaacatg aaggacaact ggcgcagcga gctgtacaag 1500
tacaaggtgg tgcgcacgca gccccctggg gtggccccca cccaggccaa gcgcgcgctg 1560
gtgcagcgcg agaagcgcg cgtgggcctg ggcgcctgt tcatcggtt cctgggccc 1620
gccgggagca ccatgggcgc cgcctcgtg accctgaccg tgcaggcccg ccagctgtg 1680
agcggcatcg tgcagcagca gaacaacctg ctgcgcgcca tcgaggccca gcagcacctg 1740
ctgcagctga cgtgtgggg catcaagcag ctgcaggccc gcacccctgg cgtggagcgc 1800
tacctgaagg accagcagct gctgggcacg tggggctgca gcggcaagct gatctgcacc 1860
accaccgtgc cctggaacag cagctggagc aacaagagcc tgaccgagat ctgggacaac 1920
atgacctgga tggagtggga gcgcgagatc ggcaactaca ccggcctgat ctacaacctg 1980
atcgagatcg cccagaacca gcaggagaag aacgagcagg agctgctgga gctggacaag 2040
tggggccagcc tgtggaactg gttcgacatc accaactggc tgtggtacat ccgcatcttc 2100
atcatgatcg tggggcgccct gatcgccctg cgcacgtgtg tcgccgtgct gagcatcgtg 2160
aaccgcgtgc gccagggcta cagccccatc agcctgcaga cccgcctgcc cgcaccagcg 2220
ggccccgacc gccccgaggc catcgaggag gaggcgggcg agcgcgaccg cgaccgcagc 2280
aaccgcctgg tgcacggcct gctggccctg atctgggacg acctgcgcag cctgtgcctg 2340
ttcagctacc accgcctgcg cgacctgctg ctgatcgtgg ccgcgcatcg ggagctgctg 2400
ggccgcgcgc gctgggaggc cctgaagtac tgggtgaacc tgctgcagta ctggagccag 2460
gagctgaaga gcagcgccgt gagcctgttc aacgccaccg ccatcgccgt ggccgagggc 2520
accgaccgca tcacgcagat cgtgcagcgc atcttcgcgc ccgtgatcca catccccgc 2580
cgcatccgcc agggcctgga gcgcgcctg ctgtaagata tcggatcctc taga 2634

```

&lt;210&gt; 65

&lt;211&gt; 2538

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

gp160.modUS4.delV1

&lt;400&gt; 65

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgtgct gctgtgtgga 60
gcagtcttcg ttctgcccag gcgccaccac gtgtgtgtgg tgaccgtgta ctacggcgtg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggcttacaag 180
gccgaggccc acaacgtgtg ggcacccac gcctgcgtgc ccaccgacc caaccccag 240
gagggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagctgg gcgccggcgg cgagatcaag 420
aactgcagct tcaacatcac caccagcgtg cgcgacaagg tgcagaagga gtacagcctg 480
ttctacaagc tggacgtggt gcccatcgac aacgacaacg ccagctaccg cctgatcaac 540
tgcaaacacca gcgtgatcac ccaggcctgc cccaaggtga gcttcgagcc catecccatc 600
cactactgcg cccccgcggg ctctgccatc ctgaagtga aggacaagaa gttcaacggc 660
accggccctt gcaagaacgt gagcaccgtg cagtgcaccc acggcatccg ccccggtgtg 720
agcaccagc tgctgtgtaa cggcagcctg gccgaggagg agatcgtgct gcgctccgag 780
aacttcaccg acaacgcca gaccatcatc gtgcagctga acgagtcctg ggagatcaac 840
tgcacccgcc ccaacaacaa cagcgtaag agcatccaca tcggccccgg ccgcgccttc 900
tacgccaccg gcgacatcat cggcgacatc cgccaggccc actgcaacat cagcaaggcc 960
aactggacca acaccctcga gcagatcgtg gagaagctgc gcgagcagtt cggcaacaac 1020
aagaccatca tcttcaacag cagcagcggc ggcgacccc agatcgtgtt ccacagcttc 1080
aactgcggcg gcgagttctt ctactgcaac accagccagc tgttcaacag cacctggaac 1140
atcaccgagg aggtgaacaa gaccaaggag aacgacacca tcacccctgcc ctgccgcatc 1200
cgccagatca tcaacatgtg gcaggagggt ggcaaggcca tgtacgcccc ccccatccgc 1260
ggccagatca agtgacagc caatattacc ggctgtgtgc tgaccgcga cggcgccacc 1320
aacaacaacc gcaccaacga caccgagacc ttccgcccc gcgccggcaa catgaaggac 1380
aactggcgca gcgagctgta caagtacaag gtggtgcgca tcgagccctt gggcgtggcc 1440
cccaccaggg ccaagcgccg cgtggtgcag cgcgagaagc gcgcctggg cctgggccc 1500
ctgttcatcg gcttcctggg cgccgcggg agcaccatgg gcgcgcctc cgtgaccctg 1560

```

```

accgtgcagg cccgccagct gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1620
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1680
ccccgcaccc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1740
tgcagcggca agctgatctg caccaccacc gtgccctgga acagcagctg gagcaacaag 1800
agcctgaccg agatctggga caacatgacc tggatggagt gggagcgcga gatcggcaac 1860
tacaccggcc tgatctacaa cctgatcgag atcgccaga accagcagga gaagaacgag 1920
caggagctgc tggagctgga caagtggggc agcctgtgga actggttcga catcaccaac 1980
tggctgtggt acatccgcat cttcatcatg atcgtgggag gctgatcgg cctgcgcac 2040
gtgttcgccc tgctgagcat cgtgaaccgc gtgcgccagg gctacagccc catcagcctg 2100
cagaccgcgc tggccgcccc gcgcggcccc gaccgccccg agggcatcga ggaggagggc 2160
ggcgagcgcg acccgagccg cagcaaccgc ctggtgcacg gcctgctggc cctgatctgg 2220
gacgacctgc gcagcctgtg cctgttcagc taccaccgcc tgcgcgacct gctgctgac 2280
gtggcccgca tctgtgagct gctggggcgc cgcggtggg agggcctgaa gtactgggtg 2340
aacctgctgc agtactggag ccaggagctg aagagcagcg ccgtgagcct gttcaacgcc 2400
accgcatcgc ccgtggccga gggcaccgac cgcacatcgc agatcgtgca gcgcacctt 2460
cgcgccgtga tccacatccc ccgccgcac cgcaggggc tggagcgcgc cctgctgtaa 2520
gatatcgcat cctctaga
2538

```

&lt;210&gt; 66

&lt;211&gt; 2553

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp160.modUS4.delV2

&lt;400&gt; 66

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg tttcgccagc cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
ccgctgtgga aggaggccac caccacctg tttcgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggcacccac gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catgggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgacctt gaactgcacc gacaagctga ccggcagcac caacggcacc 420
aacagcacca gcggcaccac cagcaccagc ggcaccaaca gcaccagcac caacagcacc 480
gacagctggg agaagatgcc cgagggcgag atcaagaact gcagcttcaa catcggcgcc 540
ggccgcctga tcaactgcaa caccagcgtg atcaccagc cctgccccaa ggtgagcttc 600
gagcccatcc ccatccacta ctgcgccccg gccggcttcg ccacctgaa gtgcaaggac 660
aagaagtcca acggcaccgg cccctgcaag aacgtgagca ccgtgcagtg caccacggc 720
atccgccccg tggtagcac ccagctgtg ctgaacggca gcctggccga ggaggagatc 780
gtgctgcgct ccgagaactt caccgacaac gccaaagacca tcatcgtgca gctgaacgag 840
tccgtggaga tcaactgcat ccgccccaac aacaacacgc gtaagagcat ccacatcggc 900
ccggcccgcg ccttctacgc caccggcgac atcatcggcg acatccgcca ggccactgc 960
aacatcagca aggccaaactg gaccaaacacc ctcgagcaga tctgaggaga gctgcgcgag 1020
cagttcgcca acaacaagac catcatcttc aacagcagca gcggcgccga ccccgagatc 1080
gtgttcacac gcttcaactg cggcgggcag ttcttctact gcaacaccag ccagctgttc 1140
aacagcacct ggaacatcac cgaggaggtg aacaagacca aggagaacga caccatcatc 1200
ctgccctgcc gcatccgcca gatcatcaac atgtggcagg aggtgggcaa ggccatgtac 1260
gcccccccca tccgcgccca gatcaagtgc agcagcaata ttaccggcct gctgctgacc 1320
cgcgacggcg gcaccaacaa caaccgcacc aacgacaccg agaccttcg ccccgcgggc 1380
ggcaacatga aggacaactg gcgcagcgag ctgtacaagt acaagggtgt gcgcacgag 1440
cccctgggag tggccccac ccaggccaag cgccgcgtgg tgcagcgcga gaagcgcgcc 1500
gtgggcctgg gcgcctgtt catcggttc ctgggcgcgc ccgggagcac catgggcgcc 1560
gcctccgtga cctgaccgt gcaggccgcg cagctgctga gcggcatcgt gcagcagcag 1620
aacaacctgc tgcgcgcat cgaggccag cagcacctgc tgcagctgac cgtgtggggc 1680
atcaagcagc tgcaggcccg catcctggcc gtggagcgt acctgaagga ccagcagctg 1740
ctgggcatct ggggctgcag cggcaagctg atctgcacca ccaccgtgcc ctggaacagc 1800
agctggagca acaagagcct gaccgagatc tgggacaaca tgacctggat ggagtgggag 1860

```

```

cgcgagatcg gcaactacac cggcctgata tacaacctga tcgagatcgc ccagaaccag 1920
caggagaaga acgagcagga gctgctggag ctggacaagt gggccagcct gtggaactgg 1980
ttcgacatca ccaactggct gtggtacatc cgcattctta tcatgatcgt gggcggcctg 2040
atcggcctgc gcatcggtgt cgcctgctg agcatcgtga accgcgtgcy ccaggggtac 2100
agcccatca gcctgcagac ccgctgccc gccagcgcg gcccgcaccg ccccgaggcg 2160
atcgaggagg agggcggcga gcgcgaccgc gaccgcagca accgcctggg gcacggcctg 2220
ctggccctga tctgggacga cctgcgcagc ctgtgcctgt tcagctacca cgcctgcgc 2280
gacctgctgc tgatcggtgg ccgcatcgtg gagctgctgg gccgcgcgcg ctgggaggcc 2340
ctgaagtact ggtggaacct gctgcagtac tggagccagg agctgaagag cagcgccctg 2400
agcctgttca acgcccacgc catcgccgtg gccaggggca ccgaccgcat catcgagatc 2460
gtgcagcgca tcttcgcgcg cgtgatccac atccccgcg gcattccgca gggcctggag 2520
cgcgcctcgc tgtaagatat cggatcctct aga

```

&lt;210&gt; 67

&lt;211&gt; 2340

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp160.modUS4.delV1/V2

&lt;400&gt; 67

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120
cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccacc gcctgcgtgc ccaccgacc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgggcgcg 360
ggccaggcct gcccgaaggt gagcttcgag cccatcccca tccactactg cgcccccgcc 420
ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
gtgagcaccg tgcagtgcac ccacggcatc cgccccgtgg tgagcaccga gctgctgctg 540
aacggcagcc tggccgagga ggagatcgtg ctgcgctccg agaacttcac cgacaacgcc 600
aagaccatca tcgtgcagct gaacgagtcg gtggagatca actgcatccg cccaacaac 660
aacacgcgta agagcatcca catcgccccc ggccgcgcct tctacgccac cggcgacatc 720
atcggcgaca tccgccaggg ccactgcaac atcagcaagg ccaactggag caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccggcaaca acaagaccat catcttcaac 840
agcagcagcg gcggcgaccc cgagatcgtg ttccacagct tcaactgcgg cggcgaggtt 900
ttctactgca acaccagcca gctgttcaac agcaccgtga acatcaccga ggaggtgaac 960
aagaccaagg agaacgacac catcatcctg ccttgccgca tccgccagat catcaacatg 1020
tggcaggagg tgggcaaggc catgtacgcc cccccatcc gcggccagat caagtgcagc 1080
agcaatatta ccggcctgct gctgaccgcg gacggcggca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttcgcgcc cggcgggcgc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgcg catcgagccc ctgggcgtgg cccccacca ggccaagcgc 1260
cgcgtggtgc agcgcgagaa gcgcgccgtg ggctggggcg cctgttctat cggttctctg 1320
ggcgccgccc ggagcaccat gggcgccgccc tccgtgaccc tgaccgtgca ggcccgccag 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgctgc gcgccatcga ggcccagcag 1440
cacctgctgc agctgaccgt gtggggcatc aagcagctgc agggccgcat cctggccgtg 1500
gagcgctacc tgaaggacca gcagctgctg ggcatctggg gctgcagcgg caagctgac 1560
tgcaccacca ccgtgccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcggca actacaccgg cctgatctac 1680
aacctgatcg agatcgccca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatccgc 1800
atcttcatca tgatcgtggg cggcctgata ggctgcgca tctgttctgc cgtgctgagc 1860
atcgtgaacc gcgtgcgcca gggctacagc cccatcagcc tgagacccg cctgcccgcg 1920
cagcgcgccc ccgaccgccc cgagggcata gcgaggagg gcggcgagcg cgaccgcgac 1980
cgcagcaacc gcctggtgca cggcctgctg gccctgatct gggacgacct gcgcagcctg 2040
tgctgttca gctaccaccg cctgcgcgac ctgctgctga tctgtggccc catcgtggag 2100
ctgctgggcc gccgcggctg ggaggccctg aagtactggg ggaacctgct gcagtactgg 2160

```

agccaggagc tgaagagcag cgccgtgagc ctgttcaacg ccaccgccat cgccgtggcc 2220  
 gagggcaccg accgcatcat cgagatcgtg cagcgcatct tccgcgcgt gatccacatc 2280  
 ccccgccgca tccgccaggg cctggagcgc gccctgctgt aagatatcgg atcctctaga 2340

<210> 68

<211> 2385

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

gp160.modUS4del 128-194

<400> 68

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60  
 gcagtcttcg tttcgccacg cgccaccacc gtgctgtggg tgaccgtgta ctacggcggtg 120  
 cccgtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcaa ggcttacaag 180  
 gccgagggccc acaacgtgtg gggcaccac gcctgctgct ccaccgaccc caaccccag 240  
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300  
 cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360  
 acccccctgt gcgtggggggc agggaaactgc gagaccagcg tgatcaccga ggccctgcccc 420  
 aaggtgagct tcgagcccat ccccatccac tactgcgccc ccgcccggctt cgccatcctg 480  
 aagtgaagg acaagaagtt caacggcacc ggcccctgca agaactgtgag caccgtgcag 540  
 tgcacccacg gcatccgccc cgtgggtgagc acccagctgc tgcgtgaacgg cagcctggcc 600  
 gaggaggaga tcgtgctgctg ctccgagaac ttaccgaca acgccaagac catcatcgtg 660  
 cagctgaacg agtccgtgga gatcaactgc atccgccccca acaacaacac gcgtaagagc 720  
 atccacatcg gcccggccg cgcttctac gccaccggcg acatcatcgg cgacatccgc 780  
 caggcccact gcaacatcag caaggccaac tggaccaaca cctcgagca gatcgtggag 840  
 aagctgcgcg agcagttcgg caacaacaag accatcatct tcaacagcag cagcggcgcc 900  
 gaccccgaga tcgtgttcca cagcttcaac tgcggcgggc agttcttcta ctgcaacacc 960  
 agccagctgt tcaacagcac ctggaacatc accgaggagg tgaacaagac caaggagaac 1020  
 gacacatca tcctgcctc cgcctccgc cagatcatca acatgtggca ggaggtgggc 1080  
 aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa tattaccggc 1140  
 ctgctgctga cccgcgacgg cggcaccaac aacaaccgca ccaacgacac cgagacctt 1200  
 cgcgccggcg gcggcaacat gaaggacaac tggcgagcgc agctgtacaa gtacaagggtg 1260  
 gtgcgcatcg agcccctggg cgtggccccc acccaggcca agcgcccgct ggtgcagcgc 1320  
 gagaagcgcg ccgtgggcct gggcgccctg ttcatcggtt tccctgggcgc cgccgggagc 1380  
 accatgggcg ccgcctccgt gaccctgacc gtgcaggccc gccagctgct gagcggcatc 1440  
 gtgcagcagc agaacaacct gctgcgcgcc atcgaggccc agcagcacct gctgcagctg 1500  
 accgtgtggg gcatcaagca gctgcaggcc cgcctcctgg ccgtggagcg ctacctgaag 1560  
 gaccagcagc tgctgggcat ctggggctgc agcggaagc tgatctgcac caccaccgtg 1620  
 ccctggaaca gcagctggag caacaagagc ctgaccgaga tctgggacaa catgacctgg 1680  
 atggagtggg agcgcgagat cggcaactac accggcctga tctacaacct gatcgagatc 1740  
 gcccagaacc agcaggagaa gaacgagcag gagctgctgg agctggacaa gtggggccagc 1800  
 ctgtggaaact ggttcgacat caccaactgg ctgtggtaca tccgcatctt catcatgatc 1860  
 gtggggcgcc tgatcgccct gcgcctcgtg ttccgctgct tgagcatcgt gaaccgctg 1920  
 cgccagggct acagccccc cagcctgcag accgcctgc ccgcccagcg cggccccgac 1980  
 cgccccgagg gcatcgagga ggaggcgcc gagcgcgacc gcgaccgag caaccgcctg 2040  
 gtgcacggcc tgctggccct gatctgggac gacctgcgca gcctgtgcct gttcagctac 2100  
 caccgcctgc gcgacctgct gctgatcgtg gcccgcatcg tggagctgct gggccgccc 2160  
 ggctgggagg ccctgaagta ctggtggaac ctgctgcagt actggagcca ggagctgaag 2220  
 agcagcgccg tgagcctggt caacgccacc gccctgcgg tggccgaggg caccgaccgc 2280  
 atcatcgaga tcgtgcagcg catcttccgc gccgtgatcc acatcccccg ccgcatccgc 2340  
 cagggcctgg agcgcgccct gctgtaagat atcggtacct ctaga 2385

<210> 69

<211> 144

<212> DNA

<213> Human immunodeficiency virus

&lt;400&gt; 69

gacaccatca tcttgccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 60  
 aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa catcaccggc 120  
 ctgctgctga cccgcgacgg cggc 144

&lt;210&gt; 70

&lt;211&gt; 144

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 70

ggaactatca cactcccatg cagaataaaa caaattataa acaggtggca ggaagtagga 60  
 aaagcaatgt atgcccctcc catcagagga caaattagat gctcatcaaa tattacagga 120  
 ctgtatttaa caagagatgg tggc 144

&lt;210&gt; 71

&lt;211&gt; 144

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic Env  
 US4 common region

&lt;400&gt; 71

gacaccatca tcttgccctg ccgcatccgc cagatcatca acatgtggca ggaggtgggc 60  
 aaggccatgt acgccccccc catccgcggc cagatcaagt gcagcagcaa catcaccggc 120  
 ctgctgctga cccgcgacgg cggc 144

&lt;210&gt; 72

&lt;211&gt; 144

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: synthetic Env  
 SF162 common region

&lt;400&gt; 72

ggcaccatca ccttgccctg ccgcatcaag cagatcatca accgctggca ggaggtgggc 60  
 aaggccatgt acgccccccc catccgcggc cagatccgct gcagcagcaa catcaccggc 120  
 ctgctgctga cccgcgacgg cggc 144

&lt;210&gt; 73

&lt;211&gt; 4766

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
 gp160.modUS4.gag.modSF2

&lt;400&gt; 73

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60  
 gcagtcttcg tttcgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120  
 cccgtgtgga aggaggccac caccacctg ttctgcgcca gcgacgcaa ggcttacaag 180  
 gccgaggccc acaacgtgtg ggccaccac gcctgctgct ccaccgaccc caacccccag 240  
 gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300  
 cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360

acccccctgt	gcgtgacct	gaactgcacc	gacaagctga	ccggcagcac	caacggcacc	420
aacagcacca	gcggcaccaa	cagcaccagc	ggcaccaaca	gcaccagcac	caacagcacc	480
gacagctggg	agaagatgcc	cgagggcgag	atcaagaact	gcagcttcaa	catcaccacc	540
agcgtgcgcg	acaaggtgca	gaaggagtac	agcctgttct	acaagctgga	cgtggtgccc	600
atcgacaacg	acaacgccag	ctaccgcctg	atcaactgca	acaccagcgt	gatcaccacc	660
gcttgcccca	agggtgagctt	cgagcccatc	cccatccact	actgcgcccc	cgccggcttc	720
gccatcctga	agtgaagga	caagaagttc	aacggcaccg	gccctgcaa	gaacgtgagc	780
accgtgcagt	gcacccaacg	catccgcocc	gtggtgagca	cccagctgct	gctgaacggc	840
agcctggccg	aggaggagat	cgtgctgcgc	tccgagaact	tcaccgacaa	cgccaagacc	900
atcatcgtgc	agctgaacga	gtccgtggag	atcaactgca	tccgccccaa	caacaacacg	960
cgtaagagca	tccacatcgg	ccccggccgc	gccttctacg	ccaccggcga	catcatcggc	1020
gacatccgcc	aggcccaactg	caacatcagc	aaggccaact	ggaccaacac	cctcgagcag	1080
atcgtggaga	agctgcgcga	gcagttcggc	aacaacaaga	ccatcatctt	caacagcagc	1140
agcggcgccg	accccgagat	cgtgttccac	agcttcaact	gcggcgccga	gttcttctac	1200
tgaacacca	gccagctgtt	caacagcacc	tggaaacatca	ccgaggaggt	gaacaagacc	1260
aaggagaacg	acaccatcat	cctgccctgc	cgcatccgcc	agatcatcaa	catgtggcag	1320
gaggtgggca	aggccatgta	cgcccccccc	atccgcgccg	agatcaagtg	cagcagcaat	1380
attaccggcc	tgctgtgcac	ccgcgacggc	ggcaccaaca	acaaccgcac	caacgcacac	1440
gagaccttcc	gccccggcgg	ggcaacatg	aaggacaact	ggcgagcga	gctgtacaag	1500
tacaaggtgg	tgcgcatcga	gccccctggg	gtggccccca	cccaggccaa	gcgcgcgctg	1560
gtgcagcgcg	agaagcgcg	cgtgggcctg	ggcgccctgt	tcatcggctt	cctgggcggc	1620
gcccgggagca	ccatgggcgc	cgctcccgctg	accctgacgg	tgcaggcccg	ccagctgctg	1680
agcggcatcg	tgcagcagca	gaacaacctg	ctgcgcgcca	tcgaggccca	gcagcacctg	1740
ctgcagctga	ccgtgtgggg	catcaagcag	ctgcaggccc	gcacctggcg	cgtggagcgc	1800
tacctgaagg	accagcagct	gctgggcacg	tggggctgca	gcggcaagct	gatctgcacc	1860
accacccgtg	cctggaacag	cagctggagc	aacaagagcc	tgaccgagat	ctgggacaac	1920
atgacctgga	tggagtggga	gcgcgagatc	ggcaactaca	ccggcctgat	ctacaacctg	1980
atcgagatcg	cccagaacca	gcaggagaag	aacgagcagg	agctgctgga	gctggacaag	2040
tggggccagcc	tgtggaactg	gttcgacatc	accaactggc	tgtggtacat	ccgcatcttc	2100
atcatgatcg	tggggggcct	gatcggcctg	cgcatcgtgt	tcgccgtgct	gagcatcgtg	2160
aaccgcgtgc	gccagggcta	cagccccatc	agcctgcaga	cccgcctgcc	cgcccagcgc	2220
ggccccgacc	gccccgaggg	catcgaggag	gagggcgccg	agcgcgaccg	cgaccgcagc	2280
aaccgcctgg	tgcacggcct	gtcggccctg	atctgggacg	acctgcgacg	cctgtgcctg	2340
ttcagctacc	accgcctgcg	cgacctgctg	ctgatcgtgg	ccgcacatcg	ggagctgctg	2400
ggccgcgcgc	gctgggaggg	cctgaagtac	tgggtggaacc	tgctgcagta	ctggagccag	2460
gagctgaaga	gcagcgccgt	gagcctgttc	aacgccaccg	ccatcgccgt	ggccgagggc	2520
accgaccgca	tcatcgagat	cgtgcagcgc	atcttccgcg	ccgtgatcca	catccccgcg	2580
cgcccccctg	agggcctgga	gcgcgcctg	ctgtaagata	tcggatcttc	tagagaattc	2640
cgcccccccc	cccccccccc	ctctccctcc	ccccccccta	acgttactgg	ccgaagccgc	2700
ttggaataag	gcccgtgtgc	gtttgtctat	atgttatttt	ccaccataat	gccgtctttt	2760
ggcaatgtga	gggcccggaa	acctggccct	gtcttcttga	cgagcattcc	taggggtctt	2820
tcccctctcg	ccaaaggaat	gcaaggtctg	ttgaatgtcg	tgaaggaaag	agttcctctg	2880
gaagcttctt	gaagacaaac	aacgtctgta	gcgacccttt	gcaggcagcg	gaacccccca	2940
cctggcgaca	ggtgcctctg	cgcccaaaag	ccacgtgtat	aagatacacc	tgcaaaggcg	3000
gcacaacccc	agtgcacagt	tgtgagttgg	atagttgtgg	aaagagtcaa	atggctctcc	3060
tcaagcgtat	tcaacaaggg	gctgaaggat	gcccagaagg	tacccatttg	tatgggatct	3120
gatctggggc	ctcgggtgcac	atgctttaca	tgtgtttagt	cgagggttaa	aaaacgtcta	3180
ggccccccga	accacgggga	cgtgggtttc	ctttgaaaaa	cagcataata	ccatgggcgc	3240
ccgcgccagc	gtgctgagcg	gcggcgagct	ggacaagtgg	gagaagatcc	gcctgcgccc	3300
cgccggcaag	aagaagtaca	agctgaagca	catcgtgtgg	gccagccgcg	agctggagcg	3360
cttcgccgtg	aaccccgccc	tgctggagac	cagcgagggc	tgccgcccga	tcctgggcca	3420
gctgcagccc	agcctgcaga	ccggcagcga	ggagctgcgc	agcctgtaca	acaccgtggc	3480
cacctgtac	tgctgtgcac	agcgcacgca	cgtcaaggac	accaaggagg	ccctggagaa	3540
gatcgaggag	gagcagaaca	agtcacaaga	gaaggcccag	caggccgccc	ccgcgcgccc	3600
caccggcaac	agcagccagg	tgagccagaa	ctaccccatc	gtgcagaacc	tgaggggcca	3660
gatggtgcac	caggccatca	gccccgcac	cctgaacgcc	tgggtgaagg	tgggtggagga	3720
gaaggccttc	agccccgagg	tgatccccat	gttcagcgcc	ctgagcgagg	gcgccacccc	3780
ccaggacatg	aacacgatgt	tgaacaccgt	gggcggccac	caggccgcca	tgcagatgct	3840
gaaggagacc	atcaacgagg	aggccgcccga	gtgggaccgc	gtgcaccccg	tgcacgcccg	3900



```

ccccatcgcc cccggccaga tgcgcgagcc ccgcggcagc gacatcgccg gcaccaccag 3960
caccctggcag gagcagatcg gctggatgac caacaacccc cccatccccg tgggcgagat 4020
ctacaagcgg tggatcatcc tgggcctgaa caagatcgtg cggatgtaca gccccaccag 4080
catcctggac atccgccagg gccccaaagg gcccttccgc gactacgtgg accgcttcta 4140
caagaccctg cgcgctgagc agggccagcca ggacgtgaag aactggatga ccgagaccct 4200
gctggtgcag aacgccaaacc ccgactgcaa gaccatcctg aaggctctcg gcccccgccg 4260
caccctggag gagatgatga ccgcctgcca gggcgtgggc ggccccggcc acaaggcccc 4320
cgtgctggcc gaggcgatga gccagggtgac gaacccggcg accatcatga tgcagcgccg 4380
caacttccgc aaccagcgga agaccgtcaa gtgcttcaac tgcggcaagg agggccacac 4440
cgccaggaac tgcgcgccc cccgcaagaa gggctgctgg cgctgcggcc gcgagggcca 4500
ccagatgaag gactgcaccg agcgccaggc caacttcctg ggcaagatct ggcccagcta 4560
caagggccgc cccggcaact tccctgcagag ccgccccgag cccaccgccc cccccgagga 4620
gagcttccgc ttcggcgagg agaagaccac ccccagccag aagcaggagc ccatcgacaa 4680
ggagctgtac cccctgacca gcctgcgcag cctgttcggc aacgacccca gcagccagta 4740
agaattcaga ctcgagcaag tctaga 4766

```

&lt;210&gt; 74

&lt;211&gt; 4689

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp160.modSF162.gag.modSF2

&lt;400&gt; 74

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgcctgagg agactgtggg tgaccgtgta ctacggcgtg 120
cccggtggga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggcctacgac 180
accgaggtgc acaacgtgtg ggccaccac gcctgcgtgc ccaccgacc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gctgaccct gcactgcacc aacctgaaga acgcccacaa caccaagagc 420
agcaactgga aggagatgga ccgcggcgag atcaagaact gcagcttcaa ggtgaccacc 480
agcatccgca acaagatgca gaaggagtac gccctgttct acaagctgga cgtggtgccc 540
atcgacaacg acaacaccag ctacaagctg atcaactgca acaccagcgt gatcaccag 600
gcctgcccc aagtgagctt cgagcccatc cccatccact actgcgccc cgccggcttc 660
gccatccgca agtgcaacga caagaagttc aacggcagcg gccctgcac caacgtgagc 720
accgtgcagt gcacccacgg catccgcccc gtggtgagca cccagctgct gctgaacggc 780
agcctggccg aggagggcgt ggtgatccgc agcgagaact tcaccgacaa cgccaagacc 840
atcatcgtgc agctgaagga gagcgtggag atcaactgca cccgccccaa caacaacacc 900
cgcaagagca tcaccatcgg ccccgccgcg gccttctacg ccaccggcga catcatcggc 960
gacatccgcc agggccactg caacatcagc ggcgagaagt ggaacaacac cctgaagcag 1020
atcgtgacca agctgcaggc ccagttcggc aacaagacca tcgtgttcaa gcagagcagc 1080
ggcgccgacc ccgagatcgt gatgcacagc ttcaactgcy gcggcgagtt cttctactgc 1140
aacagcacc agctgttcaa cagcacctgg aacaacacca tcggcccca caacaccaac 1200
ggcaccatca ccctgcccct ccgcatcaag cagatcatca accgctggca ggaggtgggc 1260
aaggccatgt acgccccccc catcccgccg cagatccgct gcagcagcaa catcaccggc 1320
ctgctgctga cccgcgacgg cggcaaggag atcagcaaca ccaccgagat cttccgcccc 1380
ggcgccggcg acatgcgcga caactggcgc agcgagctgt acaagtacaa ggtggtgaag 1440
atcgagcccc tggcgctggc ccccaaccaag gccaagcgcc gcgtggtgca gcgcgagaag 1500
cgcgccgtga ccctgggcgc catgttccctg ggcttccctg gcgcccggcg cagcaccatg 1560
ggcgcccgca gcctgaccct gaccgtgcag gcccgccagc tgctgagcgg catcgtgcag 1620
cagcagaaca acctgctgcg cgccatcgag gccagcagc acctgctgca gctgaccgtg 1680
tggggcatca agcagctgca ggcccgcgct ctggccgtgg agcgctacct gaaggaccag 1740
cagctgctgg gcactctggg ctgcagcggc aagctgatct gcaccaccgc cgtgccctgg 1800
aacgccagct ggagcaacaa gagcctggag cagatctgga acaacatgac ctggatggag 1860
tgggagcgcg agatcgacaa ctacaacca cctgatcga ggagagccag 1920
aaccagcagg agaagaacga gcaggagctg ctggagctgg acaagtgggc cagcctgtgg 1980

```

```

aactgggttcg acatcagcaa gtggctgtgg tacatcaaga ttttcatcat gatcgtgggc 2040
ggcctgggtgg gcctgcgcac cgtgttcacc gtgctgagca tegtgaaccg cgtgcgccag 2100
ggctacagcc ccctgagctt ccagaccgcg tccccgcgcc ccgcgggccg cgaccgcccc 2160
gagggcatcg aggaggagg cggcgagcgc gaccgcgacc gcagcagccc cctggtgcac 2220
ggcctgctgg ccctgatctg ggacgacctg cgcagcctgt gectgttcag ctaccaccgc 2280
ctgcgcgacc tgatcctgat cgcgcgccgc atcgtggagc tgcgtggccg ccgcggctgg 2340
gagggcctga agtactgggg caacctgctg cagtactgga tccaggagct gaagaacagc 2400
gccgtgagcc tgttcgacgc catcgccatc gccgtggccg agggcaccga ccgcacatc 2460
gaggtggccc agcgcacgag ccgcgccttc ctgcacatcc ccgcgcgcat ccgccagggc 2520
ttcgagcgcg cctgctgta actcgagcaa gtctagagaa ttcgcccccc ccccccccc 2580
ccccctctccc tcccccccc ctaacgttac ttggcgaaag cgcttggaat aaggccggtg 2640
tgcgtttgtc tatatgttat ttccaccat attgccgtct tttggcaatg tgagggcccg 2700
gaaacctggc cctgtcttct tgacgagcat tccatagggg ctttccctc tcgccaaagg 2760
aatgcaaggt ctgttgaatg tegtgaagga agcagttcct ctggaagctt cttgaagaca 2820
aacaacgtct gtacgagccc ttgcaaggca gcggaacccc ccacctggcg acaggtgcct 2880
ctgcggccaa aagccacgtg tataagatac acctgcaaa gcggcacaac ccagtgcca 2940
cgttgtgagt tggatagttg tggaaagagt caaatggctc tctcaagcg tattcaacaa 3000
ggggctgaag gatgccaga aggtaccca ttgtatggga tctgatctgg ggcctcggtg 3060
cacatgcttt acatgtgttt agtcgaggtt aaaaaaacgt ctaggcccc cgaaccacgg 3120
ggacgtgggt ttcccttgaa aaacacgata ataccatggg cgcgcgcgc agcgtgctga 3180
gcggcgccga gctggacaag tgggagaaga tccgcctgcg ccccgcggc aagaagaagt 3240
acaagctgaa gcacatcgtg tgggccagcc gcgagctgga gcgcttcgcc gtgaaccccg 3300
gcctgctgga gaccagcgag ggctgccgcc agatcctggg ccagctgcag ccagcctgc 3360
agaccggcag cgaggagctg cgcagcctgt acaacaccgt ggccaccctg tactgctg 3420
accagcgcat cgacgtcaag gacaccaagg aggcctgga gaagatcgag gaggagcaga 3480
acaagtccaa gaagaaggcc cagcaggccg ccgcgcgcgc cggcaccggc aacagcagcc 3540
aggtagccaa gaactacccc atcgtgcaga acctgcagg ccagatgggtg caccaggcca 3600
tcagcccccg caccctgaac gcctgggtga aggtgggtga ggagaaggcc ttcagccccg 3660
aggtagctcc catgttcagc gccctgagcg agggcgccac ccccaggac ctgaacacga 3720
tgtagaacac cgtggcgccg caccaggccg ccatgcagat gctgaaggag accatcaacg 3780
aggagggcgc cgagtgggac cgcgtgcacc cgcgtcacgc cggccccatc gccccggcc 3840
agatgcgcga gccccgcggc agcgacatcg ccggcaccac cagcaccctg caggagcaga 3900
tcggtctggat gaccaacaac ccccccatcc ccgtggcgga gatctacaag cgttggatca 3960
tcctgggcct gaacaagatc gtgcggatgt acagccccac cagcatcctg gacatccgcc 4020
aggcccccaa ggagcccttc cgcgactacg tggaccgctt ctacaagacc ctgcgcgctg 4080
agcaggccag ccaggacgtg aagaactgga tgaccgagac cctgctgggtg cagaacgcca 4140
accccgactg caagaccatc ctgaaggctc tcggccccgc ggccaccctg gaggagatga 4200
tgaccgcctg ccaggggcgtg ggcgcccccg gccacaaggc ccgcgtgctg gccgaggcga 4260
tgagccaggt gacgaacccc gcgaccatca tgatgcagcg cggcaacttc cgcaaccagc 4320
ggaagaccgt caagtgttc aactcgggca aggagggcca caccgcccag aactgccgcg 4380
cccccgcaa gaagggtgc tggcgctgcg gcccgaggg ccaccagatg aaggactgca 4440
ccgagcgcca ggccaacttc ctgggcaaga tctggcccag ctacaagggc cggccccgca 4500
acttcctgca gagccgcccc gagccaccg cccccccga ggagagcttc cgcttcggcg 4560
aggagaagac cacccccagc cagaagcagg agcccatcga caaggagctg taccctcga 4620
ccagcctgcg cagcctgttc ggcaacgacc ccagcagcca gtaagaattc agactcgagc 4680
aagtctaga

```

&lt;210&gt; 75

&lt;211&gt; 4472

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
gp160.modUS4.delV1/V2.gag.modSF2

&lt;400&gt; 75

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtcttcg ttctgcccag cgccaccacc gtgctgtggg tgaccgtgta ctacggcgtg 120

```

```

cccggtgtgga aggaggccac caccaccctg ttctgcgcca ggcagccaa ggcttacaag 180
gccgaggccc acaacgtgtg ggccaccac gcttgcgtgc ccaccgaccc caacccccag 240
gaggtgaacc tgaccaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcatg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgggccc 360
ggccaggcct gccccaaagt gagcttcgag cccatcccca tccactactg cgtccccgccc 420
ggcttcgcca tcctgaagtg caaggacaag aagttcaacg gcaccggccc ctgcaagaac 480
gtgagcaccg tgcagtgcac ccacggcatc cgtcccggtg tgagcaccga gctgctgctg 540
aacggcagcc tggccgagga ggagatcgtg ctgctcctcg agaacttcac cgacaacgcc 600
aagaccatca tcgtgcagct gaacgagtcg gtggagatca actgcatccg cccaacaac 660
aacacgcgta agagcatcca catcgcccc ggccgcgctt tctacgccac cggcgacatc 720
atcgccgaca tccgcaaggc ccactgcaac atcagcaagg ccaactggac caacaccctc 780
gagcagatcg tggagaagct gcgcgagcag ttccgcaaca acaagacat catcttcaac 840
agcagcagcg gcggcgaccc cgagatcgtg tccacagct tcaactgcgg cggcgagtgc 900
ttctactgca acaccagcca gctgttcaac agcacttggg acatcaccga ggaggtgaac 960
aagaccaagg agaacgacac catcatcctg ccttgcgcca tccgccagat catcaacatg 1020
tggcaggagg tgggcaaggc catgtacgcc cccccatcc gcggccagat caagtgcagc 1080
agcaatatta cggcctgct gctgacccgc gacggcgcca ccaacaacaa ccgcaccaac 1140
gacaccgaga ccttccgccc cggcgggccc aacatgaagg acaactggcg cagcgagctg 1200
tacaagtaca aggtggtgag catcgagccc ctggcggtgg cccccacca ggccaagcgc 1260
cgcgtggtgc agcgcgagaa gcgcgcgctg ggcttgggcg cctgttcat cggcttccctg 1320
ggcgccgccc ggagcaccat gggcgccgcc tccgtgaccc tgaccgtgca ggcccgcag 1380
ctgctgagcg gcatcgtgca gcagcagaac aacctgctgc gcgccatcga ggcccagcag 1440
cacctgctgc agctgacccg gtggggcatc aagcagctgc agggccgcct cctggccgctg 1500
gagcgtatcc tgaaggacca gcagctgctg gcatcctggg gctgcagcgg caagctgac 1560
tgcaccacca cgtgccctg gaacagcagc tggagcaaca agagcctgac cgagatctgg 1620
gacaacatga cctggatgga gtgggagcgc gagatcggca actacaccgg cctgatctac 1680
aacctgatcg agatcgcaca gaaccagcag gagaagaacg agcaggagct gctggagctg 1740
gacaagtggg ccagcctgtg gaactggttc gacatcacca actggctgtg gtacatccgc 1800
atcttcatca tgatcgtggg cggcctgac gcctgcgcca tctgttccgc cgtgctgagc 1860
atcgtgaacc gcgtgcgcca gggctacagc cccatcagcc tgcagacccg cctggccgcc 1920
cagcgccggc ccgaccgccc cgagggcac gaggaggagg gcggcgagcg cgaccgcgac 1980
cgcagcaacc gcctggtgca cggcctgctg gcctgatct gggacgacct gcgcagcctg 2040
tgcctgttca gctaccaccg cctgcgcgac ctgctgctga tctgtggccc catcgtggag 2100
ctgctgggccc gctgcgctg gtggccctg aagtactggg ggaacctgct gcagtactg 2160
agccaggagc tgaagagcag cgcctgagc ctgttcaacg ccaccgcat cgcctgggccc 2220
gagggcaccg accgcatcat cgagatcgtg cagcgcatct tccgcgcgt gatccacatc 2280
ccccgccga tccgccaggg cctggagcgc gcctgctgt aagatatcgg atcctctaga 2340
gaattccgcc cccccccccc cccccctct cctccccccc ccctaactg tactggccga 2400
agccgcttgg taaaggccc gtgtgcgttt gtctatatgt tatttccac catattgccg 2460
tcttttggca atgtgagggc ccggaaacct ggccctgtct tcttgacgag cattcctagg 2520
ggtctttccc ctctcgcaca aggaatgcaa ggtctgttga atgtcgtgaa ggaagcagtt 2580
cctctggaag ctctctgaag acaacaacg tctgtagcga ccttttgag gcagcggaac 2640
ccccacctg gcgacagggt cctctgcggc caaaagccac gtgtataaga tacacctgca 2700
aaggcgccac aaccccagtg ccacgttgtg agttggatag ttgtggaaag agtcaaatgg 2760
ctctcctcaa gcgtattcaa caaggggctg aaggatgccc agaaggtacc ccattgtatg 2820
ggatctgac tggggcctcg gtgcacatgc ttacatgtg tttagtcgag gttaaaaaaa 2880
cgtctaggcc ccccgaaaca cggggacgtg gttttccttt gaaaaacacg ataataccat 2940
ggcgccccc ggcagcgtgc tgagcggcgg cgagctggag aagtgggaga agatccgcct 3000
gcgccccggc ggcaagaaga agtcaagct gaagcacatc gtgtgggcca gccgcgagct 3060
ggagcgcttc gccgtgaacc ccggcctgct ggagaccagc gagggctgcc gccagatcct 3120
gggcccagct cagcccagcc tgcagaccgg cagcgaggag ctgcgcagcc tgtacaacac 3180
cgtggccacc ctgtactgcg tgcaccagcg catcgacgtc aaggacacca aggaggccct 3240
ggagaagatc gaggaggagc agaacaagtc caagaagaag gccagcagg ccgcgcgcgc 3300
cgccggcacc ggcaacagca ccagggtgag ccagaactac cccatcgtgc agaactgca 3360
gggcccagat gtgcaccagg ccacagccc ccgcaccctg aacgcctggg tgaaggtggt 3420
ggaggagaag gccttcagcc ccgaggtgat ccccatgttc agcgcctga gcgagggcgc 3480
cacccccag gacctgaaca cgatgttgaa caccgtgggc ggccaccagg ccgccatgca 3540
gatgctgaag gagaccatca acgaggaggc cgccgagtg gaccgcgtgc acccgtgca 3600
cgccggcccc atcgccccg gccagatgcg cgagccccg ggcagcgaca tcgcccgcac 3660

```

```

caccagcacc ctgcaggagc agatcggtctg gatgaccaac aaccccccca tccccgtggg 3720
cgagatctac aagcgggtgga tcatacctggg cctgaacaag atcgtgcgga tgtacagccc 3780
caccagcatc ctggacatcc gccaggggccc caaggagccc ttccgcgact acgtggaccg 3840
cttctacaag accctgcgcg ctgagcaggc cagccaggac gtgaagaact ggatgaccga 3900
gaccctgctg gtgcagaacg ccaacccccg ctgcaagacc atcctgaagg ctctcggccc 3960
cgcgggccacc ctggaggaga tgatgaccgc ctgccagggc gtgggcggcc ccggccacaa 4020
ggccccgctg ctggccgagg cgtatgagcca ggtgacgaac ccggcgacca tcatgatgca 4080
gcgcggcaac ttccgcaacc agcgggaagac cgtcaagtgc ttcaactgcg gcaaggaggg 4140
ccacaccgcc aggaactgcc gcgccccccc caagaagggc tgctggcgct gcggcccgca 4200
ggggcaccag atgaaggact gaccggagcg ccaggccaac ttccctggga agatctggcc 4260
cagctacaag ggccgccccg gcaacttcct gcagagccgc cccgagccca ccgccccccc 4320
cgaggagagc ttccgcttcg gcgaggagaa gaccaccccc agccagaagc agggagccat 4380
cgacaaggag ctgtaccccc tgaccagcct gcgcagcctg ttccggcaac accccagcag 4440
ccagtaagaa ttcagactcg agcaagtcta ga 4472

```

&lt;210&gt; 76

&lt;211&gt; 4608

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence:  
 gp160.modSF162.delV2.gag.modSF2

&lt;400&gt; 76

```

gaattcgcca ccatggatgc aatgaagaga gggctctgct gtgtgctgct gctgtgtgga 60
gcagtccttcg ttccgcccag cgccgtggag aagctgtggg tgaccgtgta ctacggcgctg 120
cccggtgtgga aggaggccac caccaccctg ttctgcgcca gcgacgcca ggccctacgac 180
accgaggtgc acaacgtgtg ggccacccac gcctgcgtgc ccaccgaccc caacccccag 240
gagatcgtgc tggagaacgt gaccgagaac ttcaacatgt ggaagaacaa catggtggag 300
cagatgcacg aggacatcat cagcctgtgg gaccagagcc tgaagccctg cgtgaagctg 360
acccccctgt gcgtgaccct gcactgcacc aacctgaaga acgccacca caccaagagc 420
agcaactgga aggagatgga ccgcgcgag atcaagaact gcagcttcaa ggtggcgcc 480
ggcaagctga tcaactgcaa caccagcgtg atcaccagg cctgccccaa ggtgagcttc 540
gagcccatcc ccatccacta ctgcgcccc gcggtcttcg ccatcctgaa gtgcaacgac 600
aagaagtcca acggcagcgg cccctgcacc aacgtgagca ccgtgcagtg caccacggc 660
atccgccccg tggtagcac ccagctgctg ctgaacggca gcctggcgca ggaggcgctg 720
gtgatccgca gcgagaactt caccgacaac gccaaagcca tcatcgtgca gctgaaggag 780
agcgtggaga tcaactgcac ccgccccaac aacaacaccc gcaagagcat caccatcggc 840
ccggcccgcg ccttctacgc caccggcgac atcatcgcg acatccgcca ggccactgc 900
aacatcagcg gcgagaagtg gaacaacacc ctgaagcaga tcgtgaccaa gctgcaggcc 960
cagttcggca acaagaccat cgtgttcaag cagagcagcg gcggcgaccc cgagatcgtg 1020
atgcacagct tcaactgcgg cggcgagtgc ttctactgca acagcaccga gctgttcaac 1080
agcacctgga acaacaccat cggccccaac aacaccaacg gcaccatcac cctgccctgc 1140
cgcatacagc agatcatcaa ccgctggcag gaggtgggca aggccatgta cggccccccc 1200
atcccgggcc agatccgctg cagcagcaac atcaccggcc tgctgctgac ccgcgacggc 1260
ggcaaggaga tcagcaaac caccgagatc ttccgcccc gcggcgcgca catgcgcgac 1320
aactggcgca gcgagctgta caagtacaag gtggtgaaga tcgagccctt gggcggtggc 1380
cccaccaagg ccaagcgcg cgtggtgcag cgcgagaagc gcgccgtgac cctggcgccc 1440
atgttcctgg gcttcctggg cgccgcccgc agcaccatgg gcgcccgag cctgaccctg 1500
accgtgcagg cccgacagc gctgagcggc atcgtgcagc agcagaacaa cctgctgcgc 1560
gccatcgagg cccagcagca cctgctgcag ctgaccgtgt ggggcatcaa gcagctgcag 1620
gcccgcgtgc tggccgtgga gcgctacctg aaggaccagc agctgctggg catctggggc 1680
tgacgcggca agctgatctg caccaccgcc gtgccctgga acgccagctg gagcaacaag 1740
agcctggacc agatctggaa caacatgacc tggatggagt gggagcgcg gatcgacaac 1800
tacaccaacc tgatctacac cctgatcgag gagagccaga accagcagga gaagaacgag 1860
caggagctgc tggagctgga caagtggcc agcctgtgga actggttcga catcagcaag 1920
tggctgtggt acatcaagat cttcatcatg atcgtggcg gcctggtggg cctgcgcac 1980
gtgttcaccg tgctgagcat cgtgaaccgc gtgcgcccag gctacagccc cctgagcttc 2040

```

```

cagacccgct tccccgcccc ccgcgcccc gaccgcccc agggcatcga ggaggagggc 2100
ggcgagcgcg acccgagccg cagcagcccc ctggtgcacg gcctgctggc cctgatctgg 2160
gacgacctgc gcagcctgtg cctgttcagc taccaccgcc tgcgcgacct gatcctgata 2220
gccgccccga tcgtggagct gctgggcccgc cgcggctggg aggccctgaa gtactggggc 2280
aacctgtctg agtactggat ccaggagctg aagaacagcg ccgtgagcct gttcgagcgc 2340
atcgccatcg ccgtggccga gggcacccgac cgcatacctg aggtggccca gcgcacggc 2400
cgcgccttcc tgcacatccc ccgcccgcac cgcaggggct tcgagcgcgc cctgctgtaa 2460
ctcgagcaag tctagagaat tccgcccccc cccccccccc cctctcctcc cccccccccc 2520
taacgttact ggccgaagcc gcttggaata aggccggtgt gcgtttgtct atatgttatt 2580
ttccaccata ttgccgtctt ttggcaatgt gagggcccg aaacctggcc ctgtcttctt 2640
gacgagcatt cctaggggtc ttcccccctc cgccaaagga atgcaaggtc tgttgaatgt 2700
cgtgaaggaa gcagttccctc tggaaagctt ttgaagacaa acaacctctg tagcgacct 2760
ttgcaggcag cggaaaccccc cacctggcga caggtgcctc tgcggccaaa agccacgtgt 2820
ataagataca cctgcaaaag cggcacacc ccagtgcac gttgtgagtt ggatagttgt 2880
ggaaagagtc aaatggctct cctcaagcgt attcaacaag gggctgaagg atgccagaa 2940
ggtagcccat tgtatgggat ctgatctggg gcctcggtgc acatgcttta catgtgttta 3000
gtcgaggtta aaaaaacgtc taggcccccc gaaccacggg gacgtgggtt tcctttgaaa 3060
aacacgataa taccatgggc gccgcgcca gcgtgctgag cggcgccgag ctggacaagt 3120
gggagaagat ccgcctgcgc cccggcgcca agaagaagta caagctgaag cacatcgtgt 3180
gggcccagcg cgagctggag cgcttcgccc tgaaccccg cctgctggag accagcgagg 3240
gctgcgcca gatcctgggc cagctgcagc ccagcctgca gaccggcagc gaggagctgc 3300
gcagcctgta caaacctgt gccacctgt actgcgtgca ccagcgcac gacgtcaagg 3360
acaccaagga ggccttgag aagatcgagg aggagcagaa caagtccaag aagaaggccc 3420
agcaggccgc ccgcgcgcc gccaccggca acagcagcca ggtgagccag aactaccca 3480
tcgtgcagaa cctgcagggc cagatggtgc accaggccat cagccccgc accctgaacg 3540
cctgggtgaa ggtggtggag gagaaggcct tcagccccga ggtgatcccc atgttcagcg 3600
ccctgagcga gggcgccacc ccccaggacc tgaacacgat gttgaacacc gtggcgggcc 3660
accaggccgc catgcagatg ctgaaggaga ccatcaacga ggaggccgc gagggggacc 3720
gcgtgcaccc cgtgcacgcc ggccccatcg ccccgccca gatgcgcgag ccccgccgca 3780
gcgacatcgc cggcaccacc agcacctgc aggagcagat cggctggatg accaacaacc 3840
cccccatccc cgtggcgag atctacaagc ggtggatcat cctgggcctg aacaagatcg 3900
tgcggtgta cagccccacc agcatcctgg acatccgcca gggccccaag gagccctcc 3960
gcgactacgt ggaccgcttc tacaagacc tgcgcgctga gcaggccagc caggacgtga 4020
agaactggat gaccgagacc ctgctggtgc agaacgcca ccccgactgc aagaccatcc 4080
tgaaggctct cggccccgcg gccacctgg aggatgatg gaccgcctgc caggggcgtg 4140
gcggccccgg ccacaaggcc cgctgctgg ccgaggcgat gagccagggt acgaacccgg 4200
cgaccatcat gatgcagcg ggcaacttc gcaaccagcg gaagaccgtc aagtgttca 4260
actgcggcaa ggagggccac accgccagga actgcgcgc cccccgcaag aagggtgct 4320
ggcgtgcgg ccgcgagggc caccagatga aggactgcac cgagcgccag gccaaacttc 4380
tgggcaagat ctggcccagc tacaagggcc gccccggcaa ctctctgcag agccgccccg 4440
agccaccgc ccccccgag gagagcttcc gcttcggcga ggagaagacc acccccagcc 4500
agaagcagga gcccatcgac aaggagctgt acccctgac cagcctgcgc agcctgttcg 4560
gcaacgaccc cagcagccag taagaattca gactcgagca agtctaga 4608

```

&lt;210&gt; 77

&lt;211&gt; 1680

&lt;212&gt; DNA

&lt;213&gt; Human immunodeficiency virus

&lt;400&gt; 77

```

cccatagtc ctattgaac tgtaccagta aaattaaagc caggaatgga tggcccaaaa 60
gttaagcaat ggccattgac agaagaaaaa ataaaagcat tagtagagat atgtacagaa 120
atggaaaagg aagggaat ttcaaaaatt gggcctgaaa atccatacaa tactccagta 180
tttgctataa agaaaaaaga cagtactaaa tggagaaaac tagtagattt cagagaactt 240
aataaaaaga ctcaagactt ctgggaagtt cagttaggaa taccacaccc cgcagggtta 300
aaaaagaaaa aatcagtaac agtattggat gtgggtgatg catacttttc agttccctta 360
gataaagact ttagaaagta tactgcattt accataccta gtataaacia tgagacacca 420
gggattagat atcagtacaa tgtgctgcca cagggtgga aaggatcacc agcaatattc 480
caaagtagca tgacaaaaat cttagagcct tttagaaaac agaattccaga catagttatc 540

```

```

tatcaatata tggatgattt gatatgtagga tctgacttag aaatagggca gcatagaaca 600
aaaatagagg aactgagaca gcatctgttg aggtggggat ttaccacacc agacaaaaaa 660
catcagaaag aacctccatt cctttggatg gggtatgaac tccatcctga taaatggaca 720
gtacagccta taatgctgcc agaaaaagac agctggactg tcaatgacat acagaagtta 780
gtgggaaaaat tgaattgggc aagtcagatt tatgcaggga ttaaagttaa gcagttatgt 840
aaactcctta gaggaacca aagcactaca gaagtaatac cactaacaga agaagcagag 900
ctagaactgg cagaaaacag ggagattcta aaagaaccag tacatgaagt atattatgac 960
ccatcaaaag acttagtagc agaaatacag aagcaggggc aaggccaatg gacatatcaa 1020
atztatcaag agccatttaa aaatctgaaa acaggaaaagt atgcaaggat gagggtgccc 1080
cacactaatg atgtaaaaa gttaacagag gcagtgcaaa aagtatccac agaaagcata 1140
gtaatatggg gaaagattcc taaatttaaa ctaccatac aaaaggaaac atgggaagca 1200
tggtggatgg agtatggca agctacctgg attctgagt gggagtttgt caataccctt 1260
cccttagtga aattatggta ccagttagag aaagaaccca tagtaggagc agaaactttc 1320
tatgtagatg gggcagctaa tagggagact aaattaggaa aagcaggata tgttactgac 1380
agaggaaagc aaaaagtgtg ctccatagct gacacaacaa atcagaagac tgaattacaa 1440
gcaattcatc tagctttgca ggattcggga ttagaagtaa acatagtaac agactcaca 1500
tatgcattag gaatcattca agcacaacca gataagagt aatcagagtt agtcagtcaa 1560
ataatagagc agttaataaa aaaggaaaag gtctacctgg catgggtacc agcacacaaa 1620
ggaattggag gaaatgaaca agtagataaa ttatgcagtg ctggaatcag gaaagtacta 1680

```

&lt;210&gt; 78

&lt;211&gt; 1865

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: GP1

&lt;400&gt; 78

```

gtcgacgcca ccattggcgc ccgcgccagc gtgctgagcg gcggcgagct ggacaagtgg 60
gagaagatcc gcctgcgccc cggcggcaag aagaagtaca agctgaagca catcgtgtgg 120
gccagccgcg agctggagcg ctctgcgctg aacccccggc tgctggagac cagcgagggc 180
tgccgccaga tcctggggcca gctgcagccc agcctgcaga ccggcagcga ggagctgcgc 240
agcctgtaca acaccgtggc caccctgtac tgcgtgcacc agcgcatcga cgtcaaggac 300
accaaggagg ccctggagaa gatcgaggag gacgagaaca agtccaagaa gaaggcccag 360
caggccgccc ccgcgcgccc caccggcaac agcagcccag tgagccagaa ctaccccatc 420
gtgcagaacc tgcaggggcca gatggtgcac caggccatca gccccgcac cctgaacgcc 480
tggttgaaag tggtggagga gaaggccttc agccccgagg tgatccccat gttcagcgcc 540
ctgagcgagg gcgccacccc ccaggacctg aacacgatgt tgaacaccgt gggcggccac 600
caggccgcca tgcagatgct gaaggagacc atcaacgagg aggcgcgcca gtgggaccgc 660
gtgcaccccc tgcacgcgcg ccccatcgcc ccgcccaga tgcgcgagcc ccgcggcagc 720
gacatcgccc gcaccaccag caccctgcag gagcagatcg gctggatgac caacaacccc 780
cccatcccc tggtgcgagat ctacaagcgg tggatcatcc tgggcctgaa caagatcgtg 840
cggatgtaca gccccaccag catcctggac atccgcccag gcccgaagga gcccttccgc 900
gactacgtgg accgcttcta caagaccctg cgcgctgagc aggcagcca ggacgtgaag 960
aactggatga ccgagaccct gctggtgcag aacgccaacc ccgactgcaa gaccatcctg 1020
aaggctctcg gccccgcggc caccctggag gagatgatga ccgcctgcca gggcgtgggc 1080
ggccccggcc acaaggcccc cgtgctggcc gaggcgatga gccaggtgac gaacccggcg 1140
accatcatga tgcagcgcg caacttccgc aaccagcgga agaccgtcaa gtgcttcaac 1200
tgccgcaagg agggccacac cgccaggaac tgccgcgccc cccgcaagaa gggctgctgg 1260
cgctgcggcc gcgaaggaca ccaaatgaaa gattgcactg agagacaggc taatttttta 1320
gggaagatct ggccttccta caagggaagg ccagggaatt ttcttcagag cagaccagag 1380
ccaacagccc caccagaaga gagcttcagg ttggggagg agaaaacaac tccctctcag 1440
aagcaggagc cgatagacaa ggaactgtat cctttaactt cctcagatc actctttggc 1500
aacgaccctt cgtcacagta aggatcggc gccagctcaa ggaggcgtg ctcgacaccg 1560
gcgcgacga caccgtgctg gaggagatga acctgcccgg caagtggag cccaagatga 1620
tcggcgggat cgggggcttc atcaaggtgc ggcagtacga ccagatcccc gtggagatct 1680
ggggccacaa ggccatcgcc accgtgctgg tgggccccac ccccgatgaac atcatcgccc 1740
gcaacctgct gaccagatc ggctgcaccc tgaacttccc catcagcccc atcgagacgg 1800

```

tgccccgtgaa gctgaagccg gggatggacg gcccccaaggt caagcagtgg cccctgtaag 1860  
aattc 1865

<210> 79

<211> 1865

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: GP2

<400> 79

```
gtcgcagcca ccatggggcgc ccgcgccagc gtgctgagcg gggcgagct ggacaagtgg 60
gagaagatcc gcctgcgccc cggcgccaag aagaagtaca agctgaagca catcgtgtgg 120
gccagcccgcg agctggagcg cttcgccgtg aacccccggc tgctggagac cagcgagggc 180
tgccgccaga tcctggggcca gctgcagccc agcctgcaga ccggcagcga ggagctgcgc 240
agcctgtaca acaccgtggc caccctgtac tgcgtgcacc agcgcatcga cgtcaaggac 300
accaaggagg ccctggagaa gatcgaggag gagcagaaca agtccaagaa gaaggcccag 360
caggcccgccg ccgcgcggcg caccggcaac agcagccagg tgagccagaa ctaccccatc 420
gtgcagaacc tgcagggcca gatggtgcac caggccatca gccccgcac cctgaacgcc 480
tgggtgaagg tgggtgagga gaaggccttc agccccgagg tgatcccatc gttcagcgcc 540
ctgagcgagg tgcccacccc ccaggacctg aacacgatgt tgaacaccgt gggcgccac 600
caggccgcca tgcagatgct gaaggagacc atcaacgagg agggcgccga gtgggaccgc 660
gtgcaccccc tgcacgcccg ccccatcgcc ccggccaga tgcgcgagcc ccgcggcagc 720
gacatcgccg gcaccaccag caccctgcag gagcagatcg gctggatgac caacaacccc 780
cccatccccg tggcgagat ctacaagcgg tggatcatcc tgggcctgaa caagatcgtg 840
cggatgtaca gccccaccag catcctggac atccgccagg gcccgaagga gcccttccgc 900
gactacgtgg accgcttcta caagaccctg cgcgctgagc agggcagcca ggacgtgaag 960
aactggatga ccgagaccct gctggtgcag aacgcccaacc ccgactgcaa gaccatcctg 1020
aaggctctcg gccccgcggc caccctggag gagatgatga ccgcctgcca gggcgtgggc 1080
ggccccggcc acaaggcccg cgtgctggcc gaggcgatga gccaggtgac gaaccggcg 1140
accatcatga tgcagcgcgg caacttccgc aaccagcggg agaccgtcaa gtgcttcaac 1200
tgccgcaagg agggccacac cggcaggaac tgcgcgccc cccgcaagaa gggctgctgg 1260
cgctgcggcc gcgaaggaca ccaaatgaaa gattgcaact agagacaggc taatttttta 1320
gggaagatct ggccttctta caagggaagg ccagggaatt ttcttcagag cagaccagag 1380
ccaacagccc caccagaaga gagcttcagg ttggggagg agaaaacaac tccctctcag 1440
aagcaggagc cgatagacaa ggaactgtat cctttaactt ccctcagatc actctttggc 1500
aacgaccctc cgtcacagta aggatcgggg ggcaactcaa ggaagcgctg ctcgatacag 1560
gagcagatga tacagtatta gaagaaatga atttgccagg aaaatggaaa ccaaaaatga 1620
taggggggat cgggggcttc atcaaggatga ggcagtaaga ccagatacct gtgaaatct 1680
gtggacataa agctataggt acagtattag taggacctac acctgtcaac ataattggaa 1740
gaaatctgtt gaccagatc ggctgcacct tgaacttccc catcagccct attgagacgg 1800
tgcccgtaag gttgaagccg gggatggacg gcccccaaggt caagcaatgg ccattgtaag 1860
aattc 1865
```

<210> 80

<211> 2305

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:

FS(+).proinact.RTopt.YM

<400> 80

```
gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatt tttttagggg 60
agatctggcc ttctacaag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
cagccccacc agaagagagc ttcaggtttg gggaggagaa aacaactccc tctcagaagc 180
aggagccgat agacaaggaa ctgtatcctt taacttccct cagatcactc tttggcaacg 240
```

```

accctcgtc acaataagga tcggggggca actcaaggaa gcgctgctcg atacaggagc 300
agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
ggggatcggg ggcttcatca aggtgaggca gtacgaccag atacctgtag aaatctgtgg 420
acataaagct ataggtacag tattagtagg acctacacct gtcaacataa ttggaagaaa 480
tctgttgacc cagatcggct gcaccttgaa ctccccatc agccctattg agacggtgcc 540
cgtgaagttg aagccgggga tggacggccc caaggtcaag caatggccat tgaccgagga 600
gaagatcaag gccctggtgg agatctgcac cgagatggag aaggagggca agatcagcaa 660
gatcggcccc gagaaccctt acaacacccc cgtgttcgcc atcaagaaga aggacagcac 720
caagtggcgc aagctggtgg acttccgcga cgtgaacaag cgcacccagg acttctggga 780
ggtgcagctg ggcattcccc accccgcccg cctgaagaag aagaagagcg tgaccgtgct 840
ggacgtgggc gacgcctact tcagcgtgcc cctggacaag gacttccgca agtacaccgc 900
cttcaccatc cccagcatca acaacgagac ccccgccatc cgctaccagt acaacgtgct 960
gccccagggc ttggaaggga gccccgccc atctccagagc agcatgacca agatcctgga 1020
gcccttccgc aagcagaacc cgcacatcgt gatctaccag gccccctgt acgtgggcag 1080
cgacctggag atcggccagc accgcaccaa gatcaggag ctgcccagc acctgctgag 1140
ctggggcttc accacccccg acaagaagca ccagaaggag ccccccttc tgtggatggg 1200
ctacgagctg caccgcgaca agtgaccgt gcagccatc atgctgccc agaaggacag 1260
ctggaccgtg aacgacatcc agaagctggt gggcaagctg aactgggcca gccagatcta 1320
cgccggcatc aaggtgaagc agctgtgcaa gctgctgcgc ggcaccaagg ccctgaccga 1380
ggtgatcccc ctgaccgagg aggcgagct ggagctggcc gagaaccgag agatcctgaa 1440
ggagcccgtg cagcaggtgt actacgaccc cagcaaggac ctggtggccg agatccagaa 1500
gcagggccag ggccagtggc cctaccagat ctaccaggag cccttcaaga acctgaagac 1560
cggcaagtac gcccgcatgc gcggcgcccc caccaacgac gtgaagcagc tgaccgaggc 1620
cgtgcagaag gctgagcacc agagcatcgt gatctggggc aagatcccca agttcaagct 1680
gccccatccg aaggagacct gggaggcctg gtggatggag tactggcagg ccacctggat 1740
ccccgagtg gaggctcgtga acaccccccc cctggtgaag ctgtggtacc agctggagaa 1800
ggagcccac gtggcgccc agaccttcta cgtggacggc gccgccaacc gcgagaccaa 1860
gctgggcaag gccggtacg tgaccgaccg gggccggcag aaggtggtga gcacgcgca 1920
caccaccaac cagaagaccg agctgcaggc catccacctg gccctgcagg acagcggcct 1980
ggaggtgaac atcgtgaccg acagccagta cgccctgggc atcatccagg cccagcccga 2040
caagagcgag agcagctgg tgagccagat catcgagcag ctgatcaaga aggagaaggt 2100
gtacctggcc tgggtgccc cccacaaggg catcgccggc aacgagcagg tggacaagct 2160
ggtgacgccc ggcattccga aggtgctgtt cctgaacggc atcgatggcg gcacgtgat 2220
ctaccagtac atggacgacc tgtacgtggg cagcggcgcc cctaggatcg attaaaagct 2280
tccccgggct agcaccgggtg aattc 2305

```

&lt;210&gt; 81

&lt;211&gt; 2299

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

FS(+) .proinact.RTopt.YMWM

&lt;400&gt; 81

```

gcggccgca aggacaccaa atgaaagatt gcactgagag acaggctaatt tttttaggga 60
agatctggcc ttctacaag ggaaggccag ggaattttct tcagagcaga ccagagccaa 120
cagccccacc agaagagagc ttcagggttg gggaggagaa aacaactccc tctcagaagc 180
aggagccgat agacaaggaa ctgtatcctt taacttccct cagatcactc tttggcaacg 240
accctcgtc acaataagga tcggggggca actcaaggaa gcgctgctcg atacaggagc 300
agatgataca gtattagaag aaatgaattt gccaggaaaa tggaaaccaa aaatgatagg 360
ggggatcggg ggcttcatca aggtgaggca gtacgaccag atacctgtag aaatctgtgg 420
acataaagct ataggtagg tattagtagg acctacacct gtcaacataa ttggaagaaa 480
tctgttgacc cagatcggct gcaccttgaa ctccccatc agccctattg agacggtgcc 540
cgtgaagttg aagccgggga tggacggccc caaggtcaag caatggccat tgaccgagga 600
gaagatcaag gccctggtgg agatctgcac cgagatggag aaggagggca agatcagcaa 660
gatcggcccc gagaaccctt acaacacccc cgtgttcgcc atcaagaaga aggacagcac 720
caagtggcgc aagctggtgg acttccgcga gctgaacaag cgcacccagg acttctggga 780

```



```

gggtgcagctg ggcattccccc accccgcgcg cctgaagaag aagaagagcg tgaccgtgct 840
ggacgtgggc gacgcctact tcagcgtgcc cctggacaag gacttccgca agtacaccgc 900
cttcaccatc cccagcatca acaacgagac ccccgcatc cgctaccagt acaacgtgct 960
gccccagggc tgggaaggga gccccgccat cttccagagc agcatgacca agatcctgga 1020
gcccctccgc aagcagaacc ccgacatcgt gatctaccag gccccctgt acgtgggcag 1080
cgacctggag atcggccagc accgcaccaa gatcgaggag ctgcgccagc acctgctgcg 1140
ctggggcttc accacccccg acaagaagca ccagaaggag ccccccttc tgcccatcga 1200
gctgcacccc gacaagtggg ccgtgcagcc catcatgctg cccgagaagg acagctggac 1260
cgtgaacgac atccagaagc tgggtgggcaa gctgaactgg gccagccaga tctacgccgg 1320
catcaagggtg aagcagctgt gcaagctgct gcgcggcacc aaggccctga ccgaggtgat 1380
ccccctgacc gaggaggccg agctggagct ggccgagaac cgcgagatcc tgaaggagcc 1440
cgtgcacgag gtgtactacg accccagcaa ggacctggtg gccgagatcc agaagcaggg 1500
ccaggggccag tggacctacc agatctacca ggagcccttc aagaacctga agaccggcaa 1560
gtacgcccgc atgcgcggcg cccacaccaa cgacgtgaag cagctgaccg aggccgtgca 1620
gaaggtgagc accgagagca tcgtgatctg gggcaagatc cccaagttca agctgcccac 1680
ccagaaggag acctgggagg cctggtggat ggagtactgg caggccacct ggatccccga 1740
gtgggagttc gtgaacaccc cccccctggt gaagctgtgg taccagctgg agaaggagcc 1800
catcgtgggc gccgagacct tctacgtgga cggcgccgcc aaccgcgaga ccaagctggg 1860
caaggccggc tacgtgaccg accggggccg gcagaagggt gtgagcatcg ccgacaccac 1920
caaccagaag accgagctgc aggccatcca cctggccctg caggacagcg gcctggaggt 1980
gaacatcgtg accgacagcc agtacgccct gggcatcatc caggcccagc ccgacaagag 2040
cgagagcgag ctgggtgagc agatcatcga gcagctgata aagaaggaga aggtgtacct 2100
ggcctgggtg ccgcccaca agggcatcgg cggcaacgag caggtggaca agctgggtgag 2160
cgccggcatc cgcaagggtg tgttctgaa cggcatcgat ggcgccatcg tgatctacca 2220
gtacatggac gacctgtacg tgggcagcgg cggccctagg atcgattaaa agcttcccgg 2280
ggctagcacc ggtgaattc                                     2299

```

&lt;210&gt; 82

&lt;211&gt; 2306

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

FS (-).protmod.RTopt.YM

&lt;400&gt; 82

```

gcggccgcga aggacaccaa atgaaagatt gcaactgagag acaggctaata ttcttccgcg 60
aggacctggc cttcctgcag ggcaaggccc gcgagttcag cagcgagcag acccgcgcca 120
acagcccccac ccgcccgcgag ctgcaggtgt ggggcggcga gaacaacagc ctgagcgagg 180
ccggcgccga ccgccagggc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
gccccctggt gaccatcagg atcggcgccc agctcaagga ggcgctgctc gacaccggcg 300
ccgacgacac cgtgctggag gagatgaacc tgcccggcaa gtggaagccc aagatgatcg 360
gcgggatcgg gggcttcatc aagggtgcggc agtacgacca gatccccgtg gagatctgcg 420
gccacaaggc catcggcacc gtgctggtgg gccccacccc cgtgaacatc atcggccgca 480
acctgctgac ccagatcggc tgcaccctga acttccccat cagccccatc gagacggtgc 540
ccgtgaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
agaagatcaa ggccctggtg gagatctgca ccgagatgga gaaggaggcg aagatcagca 660
agatcggccc cgagaacccc tacaacaccc ccgtgttcgc catcaagaag aaggacagca 720
ccaagtggcg caagctggtg gacttccgcg agctgaacaa gcgcacccag gacttctggg 780
aggtgcagct gggcatcccc cccccgcgg gcctgaagaa gaagaagagc gtgaccgtgc 840
tggacgtggg cgagcctac ttcagcgtgc ccttggaaca ggacttccgc aagtacaccg 900
ccttcacatc cccagcatc aacaacgaga cccccggcat ccgctaccag tacaacgtgc 960
tgccccaggg ctggaagggc agccccgcca tcttccagag cagcatgacc aagatcctgg 1020
agcccttccg caagcagaac cccgacatcg tgatctacca ggccccctgt tacgtgggca 1080
gcgacctgga gatcggccag caccgcacca agatcgagga gctgcgccag cacctgctgc 1140
gctggggctt caccaccccc gacaagaagc accagaagga gcccccttc ctgtggatgg 1200
gctacgagct gcaccccgac aagtggaccg tcgagcccat catgctgccc gagaaggaca 1260
gctggaccgt gaacgacatc cagaagctgg tgggcaagct gaactgggcc agccagatct 1320

```

```

acgccggcat caaggtgaag cagctgtgca agctgctgcg cggcaccaag gccctgaccg 1380
aggtgatccc cctgaccgag gaggccgagc tggagctggc cgagaaccgc gagatcctga 1440
aggagcccgt gcacgaggtg tactacgacc ccagcaagga cctggtggcc gagatccaga 1500
agcagggcca gggccagtgg acctaccaga tctaccagga gcccttcaag aacctgaaga 1560
ccggcaagta cggccgcatg cgcggcgccc acaccaacga cgtgaagcag ctgaccgagg 1620
ccgtgcagaa ggtgagcacc gagagcatcg tgatctgggg caagatcccc aagttcaagc 1680
tgccccatcca gaaggagacc tgggaggcct ggtggatgga gtactggcag gccacctgga 1740
tccccgagtg ggagttcgtg aacaccccc ccctggtgaa gctgtggtac cagctggaga 1800
aggagcccat cgtgggcgcc gagaccttct acgtggacgg cggcgccaac cgcgagacca 1860
agctgggcaa ggccggctac gtgaccgacc gggggccggca gaaggtggtg agcatcgccg 1920
acaccaccaa ccagaagacc gagctgcagg ccattccact ggccctgcag gacagcggcc 1980
tggaggtgaa catcgtgacc gacagccagt acgccctggg catcatccag gccagcccg 2040
acaagagcga gagcagctg gtgagccaga tcatcgagca gctgatcaag aaggagaagg 2100
tgtacctggc ttgggtgccc gccacaagg gcatcgccgg caacgagcag gtggacaagc 2160
tggtagcgcc cggcatccgc aaggtgctgt tcctgaacgg catcgatggc ggcatcgtga 2220
tctaccagta catggacgac ctgtacgtgg gcagcggcgg ccctaggatc gattaaaagc 2280
ttccccgggc tagcaccggt gaattc 2306

```

&lt;210&gt; 83

&lt;211&gt; 2300

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

FS(-).protmod.RTopt.YMWM

&lt;400&gt; 83

```

gcggccgcga aggacaccaa atgaaagatt gcactgagag acaggctaatt ttcttccgcg 60
aggacctggc cttcctgcag ggcaaggccc gcgagttcag cagcgagcag acccgcgcca 120
acagccccac ccgcccgcgag ctgcagggtgt ggggcggcga gaacaacagc ctgagcgagg 180
ccggcgccga ccgccagggc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
gccccctggt gaccatcagg atcggcggcc agctcaagga ggcgctgctc gacaccggcg 300
ccgacgacac cgtgctggag gagatgaacc tgcccggcaa gtggaagccc aagatgatcg 360
gcgggacceg gggcttcacg aaggtgcggc agtacgacca gatccccgtg gagatctgcg 420
gccacaaggc catcggcacc gtgctggtgg gccccacccc cgtgaacatc atcggccgca 480
acctgctgac ccagatcgcc tgcacctga acttccccat cagccccatc gagacggtgc 540
ccgtgaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
agaagatcaa ggccctgggt gagatctgca ccgagatgga gaaggaggcg aagatcagca 660
agatcgcccc cgagaacccc tacaacaccc ccgtgttcgc catcaagaag aaggacagca 720
ccaagtggcg caagctgggt gacttccgcg agctgaacaa gcgcacccag gacttctggg 780
aggtgcagct gggcatcccc cccccgcgg cctgaagaa gaagaagagc gtgaccgtgc 840
tggacgtggg cgacgcctac ttcagcgtgc ccctggcaaa ggacttccgc aagtacaccg 900
ccttcacat cccagcatc aacaacgaga cccccggcat ccgctaccag tacaacgtgc 960
tgccccaggg ctggaagggg agccccgcca tcttccagag cagcatgacc aagatcctgg 1020
agcccttccg caagcagaac ccgcacatcg tgatctacca ggccccctg tacgtgggca 1080
gcgacctgga gatcggccag caccgaccca agatcgagga gctgcgccag cacctgctgc 1140
gctggggcct caccaccccc gacaagaagc accagaagga gcccccttc ctgcccacgc 1200
agctgcaccc cgacaagtgg accgtgcagc ccatcatgct gcccgagaag gacagctgga 1260
ccgtgaacga catccagaag ctgggtgggca agctgaactg ggccagccag atctacgccg 1320
gcatcaaggt gaagcagctg tgcaagctgc tgcgcggcac caaggccctg accgaggtga 1380
tccccctgac cgaggaggcc gagctggagc tggccgagaa ccgcgagatc ctgaaggagc 1440
ccgtgcacga ggtgtactac gaccccagca aggacctggt ggccgagatc cagaagcagg 1500
gccagggcca gtggacctac cagatctacc aggagccctt caagaacctg aagaccggca 1560
agtacgccc catgcgcggc gccacacca acgacgtgaa gcagctgacc gaggccgtgc 1620
agaaggtgag caccgagagc atcgtgatct ggggcaagat ccccaagttc aagctgcccc 1680
tccagaagga gacctgggag gcctgggtgga tggagtactg gcaggccacc tggatccccg 1740
agtgggagtt cgtgaacacc cccccctgg tgaagctgtg gtaccagctg gagaaggagc 1800
ccatcgtggg cgcgagacc ttctacgtgg acggcgccgc caaccgcgag accaagctgg 1860

```

```

gcaaggccgg ctacgtgacc gaccggggcc ggcagaaggt ggtgagcatc gccgacacca 1920
ccaaccagaa gaccgagctg caggccatcc acctggccct gcaggacagc ggccctggagg 1980
tgaacatcgt gaccgacagc cagtacgccc tgggcatcat ccaggccag cccgacaaga 2040
gcgagagcga gctggtgagc cagatcatcg agcagctgat caagaaggag aaggtgtacc 2100
tggcctgggt gcccgcccac aagggtatcg gcggaacga gcaggtggac aagctggtga 2160
gcgcgggcat ccgcaagggt ctgttctga acggcatcga tggcggcato gtgatctacc 2220
agtacatgga cgacctgtac gtgggcagcg gcggccctag gatcgattaa aagcttccc 2280
gggctagcac cgggtgaattc                                     2300

```

&lt;210&gt; 84

&lt;211&gt; 2312

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

FS (-).protmod.RTopt (+)

&lt;400&gt; 84

```

gcggcccgga aggacaccaa atgaaagatt gcaactgagag acaggctaatt ttcttccg 60
aggacctggc cttcctgcag ggcaaggccc gcgagttcag cagcgagcag acccgcgcca 120
acagccccac ccgcccggag ctgcagggtgt gggcgggcga gaacaacagc ctgagcgagg 180
ccggcgccga ccgcccaggc accgtgagct tcaacttccc ccagatcacc ctgtggcagc 240
gccccctggt gaccatcagg atcgcgggcc agctcaagga ggcgctgctc gacaccggcg 300
ccgacgacac cgtgctggag gagatgaacc tgcccggcaa gtggaagccc aagatgatcg 360
gcgggatcgg gggcttcac aaggtgcggc agtacgacca gatccccgtg gagatctg 420
gccacaaggc catcggcacc gtgctggtgg gccccacccc cgtgaacatc atcgccgca 480
acctgctgac ccagatcggc tgcacctga acttccccat cagccccatc gagacgggtg 540
ccgtgaagct gaagccgggg atggacggcc ccaaggtcaa gcagtggccc ctgaccgagg 600
agaagatcaa ggccctggtg gagatctgca ccgagatgga gaaggagggc aagatcagca 660
agatcggccc cgagaacccc tacaacaccc ccgtgttcgc catcaagaag aaggacagca 720
ccaagtggcg caagctggtg gacttccgcg agctgaacaa gcgcacccag gacttctggg 780
aggtgagct gggcatcccc caccgcggcg gcctgaagaa gaagaagagc gtgaccgtgc 840
tggacgtggg cgacgcctac ttcagcgtgc ccctggacaa ggacttccgc aagtacaccg 900
ccttcacat cccagcacc aacaacgaga cccccggcat ccgctaccag tacaacgtgc 960
tgccccaggg ctggaagggc agccccgcca tcttcagag cagcatgacc aagatcctgg 1020
agcccttccg caagcagaac cccgacatcg tgatctacca gtacatggac gacctgtacg 1080
tgggcagcga cctggagatc ggccagcacc gcaccaagat cgaggagctg cgccagcacc 1140
tgctgcgctg gggcttcacc acccccgaca agaagcacca gaaggagccc cccttccgtg 1200
ggatgggcta cgagctgcac cccgacaagt ggaccgtgca gcccacatg ctgcccgaga 1260
aggacagctg gacctgaaac gacatccaga agctggtggg caagctgaac tgggcccagg 1320
agatctacgc cggcatcaag gtgaagcagc tgtgcaagct gctgcggcggc accaaggccc 1380
tgaccgaggt gatccccctg accgaggagg ccgagctgga gctggccgag aaccgcgaga 1440
tctgaagga gcccgtgcac gaggtgtact acgaccccag caaggacctg gtggccgaga 1500
tccagaagca gggccagggc cagtggacct accagatcta ccaggagccc ttcaagaacc 1560
tgaagaccgg caagtacgcc cgcatgcgcg gcgcccacac caacgacgtg aagcagctga 1620
ccgaggccgt gcagaagggt agcaccgaga gcatcgtgat ctggggcaag atccccaa 1680
tcaagctgcc catccagaag gagacctggg aggcctggtg gatggagtac tggcaggcca 1740
cctggatccc cgagtgggag ttcgtgaaca cccccccct ggtgaagctg tgggtaccag 1800
tggagaagga gcccatcgtg ggcgcccaga cttctacgt ggacggcgcc gccaacccg 1860
agaccaagct gggcaaggcc ggctacgtga ccgaccgggg ccggcagaag gtggtgagca 1920
tcgcccagac caccaaccag aagaccgagc tgcaggccat ccacctggcc ctgcaggaca 1980
gcggcctgga ggtgaacatc gtgaccgaca gccagtacgc cctgggcac atccaggccc 2040
agcccgacaa gagcgagagc gagctggtga gccagatcat cgagcagctg atcaagaagg 2100
agaaggtgta cctggcctgg gtgcccgcac acaagggtat cggcggcaac gagcaggtg 2160
acaagctggg gagcgccggc atccgcaagg tgcgtttcct gaacggcatc gatggcgga 2220
tcgtgatcta ccagtacatg gacgacctgt acgtgggcag cggcggccct aggatcgatt 2280
aaaagcttcc cggggctagc accggtgaat tc                                     2312

```

<210> 85  
 <211> 306  
 <212> DNA  
 <213> Human immunodeficiency virus

<400> 85  
 atggagccag tagatcctag attagagccc tggaagcatt caggaagtca gcctaagact 60  
 gcttgtacaa attgctattg taaaaagtgt tgctttcatt gccaaagttg ttccataaca 120  
 aaaggccttag gcatctccta tggcaggaag aagcggagac agcgacgaag agctcctcca 180  
 gacagtggagg ttcatcaagt ttctctacca aagcaacccg cttcccagcc ccaagggggac 240  
 ccgacaggcc cgaaggaatc gaagaagaag gtggagagag agacagagac agatccagtc 300  
 cattag 306

<210> 86  
 <211> 101  
 <212> PRT  
 <213> Human immunodeficiency virus

<400> 86  
 Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser  
 1 5 10 15  
 Gln Pro Lys Thr Ala Cys Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe  
 20 25 30  
 His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly  
 35 40 45  
 Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser Glu Val  
 50 55 60  
 His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp  
 65 70 75 80  
 Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu  
 85 90 95  
 Thr Asp Pro Val His  
 100

<210> 87  
 <211> 306  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: tat.SF162.opt

<400> 87  
 atggagcccg tggacccccg cctggagccc tggaagcacc ccggcagcca gcccaagacc 60  
 gcctgcacca actgctactg caagaagtgc tgcttcact gccaggtgtg cttcatcacc 120  
 aagggcctgg gcatcagcta cggccgcaag aagcggcgcg agcggcgccg cgcccccccc 180  
 gacagcgagg tgcaccaggt gagcctgccc aagcagcccg ccagccagcc ccagggcgac 240  
 cccaccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgaccccggtg 300  
 cactag 306

<210> 88  
 <211> 306

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

tat.cys22.SF162.opt

&lt;400&gt; 88

```

atggagcccc tggacccccg cctggagccc tggaagcacc ccggcagcca gccaagacc 60
gccggcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcgccgcc agcgccgccg cgcccccccc 180
gacagcgagg tgcaccaggt gagcctgccc aagcagcccc ccagccagcc ccagggcgac 240
ccacccggcc ccaaggagag caagaagaag gtggagcgcg agaccgagac cgaccccggtg 300
cactag                                           306

```

&lt;210&gt; 89

&lt;211&gt; 168

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence:

tatamino.SF162.opt

&lt;400&gt; 89

```

atggagcccc tggacccccg cctggagccc tggaagcacc ccggcagcca gccaagacc 60
gcctgcacca actgctactg caagaagtgc tgcttccact gccaggtgtg cttcatcacc 120
aagggcctgg gcatcagcta cggccgcaag aagcgccgcc agcgccgcg 168

```

&lt;210&gt; 90

&lt;211&gt; 102

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: tat cys22  
SF162 protein

&lt;400&gt; 90

```

Met Glu Pro Val Asp Pro Arg Leu Glu Pro Trp Lys His Pro Gly Ser
  1           5           10          15

Gln Pro Lys Thr Ala Gly Thr Asn Cys Tyr Cys Lys Lys Cys Cys Phe
  20          25          30

His Cys Gln Val Cys Phe Ile Thr Lys Gly Leu Gly Ile Ser Tyr Gly
  35          40          45

Arg Lys Lys Arg Arg Gln Arg Arg Arg Ala Pro Pro Asp Ser Glu Val
  50          55          60

His Gln Val Ser Leu Pro Lys Gln Pro Ala Ser Gln Pro Gln Gly Asp
  65          70          75          80

Pro Thr Gly Pro Lys Glu Ser Lys Lys Lys Val Glu Arg Glu Thr Glu
  85          90          95

Thr Asp Pro Val His Glx
 100

```

